ICAL ROOM
L LIBRARY
OF MICH.

FEB 13 1939

VOL. XXIX DECEMBER, 1938 NO. 9

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

Title-page and Index to Volume XXIX (1938) . . . . . i–viii

The Significance of IQ's on the Revised Stanford-Binet Scale . . . 641
MAUD A. MERRILL

A Factor Analysis of Test-retest Performance on Attitude and Adjustment Tests. 652
WALTER J. MCNAMARA AND JOHN G. DARLEY

The Effect of Teacher Popularity upon Attitude toward School Subjects. 665
STEPHEN M. COREY AND GEORGE S. BEERY

A Critical Comparison between the Solutions of the Factor Problem Offered by Spearman and Thurstone. 671
HATTIE BAITY HEDMAN

An Evaluation of a Scout Leaders Training Program. 686
HARRY K. EBY AND FORREST A. KINGSBURY

The Calculation of Partial and Multiple Coefficients of Regression and Correlation 695
F. L. WREN

Two Methods of Estimating Beta Weights 701
ROBERT J. WHERRY

Book Review. 710

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.
BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Established 1910

EDITORS

JACK W. DUNLAP
University of Rochester

HAROLD E. JONES
University of California

PERCIVAL M. SYMONDS
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Publishers.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD.

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIX December, 1938 Number 9

THE SIGNIFICANCE OF IQ'S ON THE REVISED STANFORD-BINET SCALES*

MAUD A. MERRILL

Stanford University

The problem of interpreting IQ's obtained on the revised Stanford-Binet scales is the problem of attaching diagnostic terms to measurable differences in intelligence, a task fraught with many difficulties. As clinicians, parents, or teachers, we want to know for very practical reasons what kind of behavior to expect of a child with an IQ of 70, an IQ of 100, or one of 150. The difficulties of our task have been increased by the fact that so much has come to be expected of the IQ as an index of developmental status or else so little because it is no better than it is.

One of the difficulties of interpreting IQ's obtained on the old scale resulted from the shortness of the scale at the upper end, making it impossible to secure adequate ratings for bright subjects much over twelve or fourteen years of age because the scale did not have enough top. So, too, at the lower end of the scale there were too few tests and they were spaced at intervals too widely separated to afford an adequate measure of mental development in terms of mental age and IQ. The passing or failing of a single test deflected the score by a too large amount. Thus the same IQ did not have quite the same diagnostic significance at all points of the scale. Since the new scales remedy some of these faults of the old scale, it is necessary to reexamine our diagnostic terminology as it applies to IQ classifications determined with reference to the revised scales.

One of the factors which has entered into the determination of the boundaries between the various current IQ classifications has been the frequency of occurrence of IQ's of a given size among unselected

* I wish to go on record as being in complete agreement with Dr. Merrill's excellent discussion of this problem. Lewis M. Terman.

children. Since the most widely used classifications have been based upon the suggestions made by Terman (4, 5) in 1916 as a result of his analysis of the percentage distributions of nine hundred five IQ's obtained on the old scale, a similar distribution for the revised scales should be useful for purposes of evaluating $L - M$ IQ's in terms of those obtained on the old scale.

TABLE I.—PERCENTAGE DISTRIBUTIONS OF IQ'S ON THE OLD AND NEW SCALES

Percentile	IQ old scale	IQ revised scales		
		Total	Ages 8–12	Ages 14–18
99	130	143	143	141
98	128	138	138	137
97	125	134	135	134
95	122	131	131	131
90	116	124	124	125
85	113	121	121	121
80	110	117	118	118
75	108	114	115	115
66.6	106	111	111	111
33.3	95	96	96	96
25	92	93	92	92
20	91	90	90	90
15	88	87	86	86
10	85	83	82	82
5	78	77	76	76
3	76	73	72	72
2	73	70	70	68
1	70	65	66	62
N	905	2,904	1,016	621

A comparison of these empirical findings with the values calculated by Bernreuter and Carr (2, p. 313) on purely theoretical grounds affords a valuable check on the validity of the score equivalents. Bernreuter and Carr, using the standard score technique, have computed from the respective standard deviations of the two scales equivalent IQ's for the revised scales and the old Stanford-Binet scale. The IQ differences between the scales reported by these authors agree with our findings so far as direction of change is concerned, but differ in magnitude. Both comparisons indicate that IQ's above 100 are higher on the revised scales than on the old scale

and that IQ's below 90 are lower on the revised scales. The results of certain retests to be reported in this paper tend to substantiate these findings on direction of change.

The magnitude of change for IQ's below 90 is greater when calculated by the standard score method than has been found to be the case empirically, but for IQ's above 100 the reverse is true. However, the amount of difference is of less consequence for IQ's above 100 than for IQ's below average on both practical and theoretical grounds. The probable error of an IQ of 130 and over on either of the new scales is 3.5, whereas for IQ's below 70 it is only 1.5 IQ points (6, p. 46). The practical importance for clinicians of reliable measurements of developmental status, particularly at the lower end of the scale, is self-evident.

Because of the diagnostic importance that has been attached to IQ's at the lower end of the scale, it is important for our purpose to see what percentage of a group of approximately three thousand representative American school children falls below a given IQ on the revised scales. The lowest two per cent of children upon whom the standardization was based test 70 or below on the revised scales. Comparable figures for the old scale yielded an IQ of 73 for the group on whom that scale was standardized in terms of IQ's obtained on the old scale. The equivalent values given by Bernreuter and Carr are 66–68 and, according to their Table I, the lowest three per cent of an unselected population should reach IQ 70 or below. In view of the estimated percentage of feebleminded in the general population (1, 3, 7), these differences are highly important.

Widespread familiarity with the meaning and limitations of the old scale IQ's is our only justification for examining the new scales in the light of comparisons with results given by the old. It should be borne in mind that such comparisons are not made for purposes of validation of the revised scales, but only as an aid in the interpretation of results. Since equivalent scores are not available for *L* and *M* and the old scale for an adequate sampling of our standardization group, it has been necessary to base our further conclusions on such clinical findings as have been accumulated since the publication of the revised scales in 1937.

During the past year, one thousand five hundred seventeen children in the elementary schools of Kansas City, Mo.,* who had previously

* These test scores were made available through the courtesy of Miss Lexie Strachan, psychologist in the Public Schools of Kansas City, Mo. The authors of

been examined with the old scale, were reexamined with the revised scale. These reexaminations were made at the teacher's request when a child's school work seemed to the teacher to be inconsistent with a previously obtained IQ (either too high or too low), when some problem of school adjustment was involved, or when a child was about to be sent on to high school and his classification was in question. The group is one which might reasonably be expected to represent the extreme IQ deviations between test and retest.

The principal results of these retests are summarized in Table II.

TABLE II.—RESULTS OF RETESTS OF FIFTEEN HUNDRED SEVENTEEN SCHOOL CHILDREN

CA* at first examination	*N*	Mean IQ old scale	Mean IQ revised scale	σ_{S-B} IQ	σ_L IQ	D/σ_D of means	r_{S-B} vs. *L* IQ	Mean interval in years between tests	σ interval in years
5	149	104.00	97.25	14.20	15.60	3.90	.64	5.16	1.94
6	806	98.00	94.60	13.25	15.95	4.66	.70	5.42	1.88
7	154	89.10	86.15	13.05	15.00	1.84	.76	5.04	1.86
8	122	89.20	88.70	11.95	15.60	0.28	.75	4.59	1.77
9	114	87.04	86.34	13.27	18.68	0.33	.80	4.03	1.62
10	93	84.25	83.90	10.50	14.30	0.19	.75	3.66	1.53
11	79	91.55	85.15	10.40	15.40	3.06	.77	3.01	1.12
Av.	...	91.88	88.87	12.37	15.79	....	.74	4.42	

* CA 5 = 4-6 to 5-5

The most outstanding characteristic of these findings is the close similarity of the mean ratings obtained on the two scales. The mean IQ's differ by less than one point at three age levels, by about three points at two levels, and by more than six points at two. Of these differences three only are statistically significant, the most significant difference being found at age six because there we find the largest number of cases and the widest range of IQ's. The obtained absolute difference, however, is greater at the higher age levels.

It is apparent, also, that at the various age levels the standard deviations for the revised scales agree closely with the standard devia-

the scales are deeply indebted to Miss Strachan and her co-workers for their generous coöperation in our efforts to evaluate the scales in clinical practice.

tion which we found for the standardization group (6, p. 40). The average IQ variability for our Kansas City data is 15.8 points for the revised scale and 12.4 points for the old scale.

Test retest correlations range from .64 to .80. The mean interval between testings is nearly four and a half years, slightly greater for the group of children who were examined first at five, six, and seven years of age and slightly less for those examined first at nine, ten, and eleven.

In view of the basis of selection of the cases to be retested—*i.e.*, discrepancy between the originally obtained IQ and the character of the child's school performance—large individual gains and losses were to be expected. Table III shows a distribution of the individual gains and losses for the one thousand five hundred seventeen children of the Kansas City study.

In spite of the fact that the results summarized in this table are based on a very biased sampling of the Kansas City school population, representing a restricted range of talent due to selective factors determining the choice of children to be reexamined, the data of the table afford empirical verification of the *direction* of IQ change indicated by Bernreuter and Carr and by our percentage distributions of IQ's on the two scales (Table I). Table III shows by how many points the *L* IQ of a given case deviates from the IQ previously obtained for that case on the old scale. An inspection of the distribution of gains and losses indicates that IQ losses were found more frequently in cases testing below 100 and IQ gains more frequently in cases testing above 100. The absolute amounts of loss and gain are not important because of the selective factors before mentioned that limited the choice of cases for reexamination. These data concern changes that affect the lower age ranges only, and there is no indication of the direction of change for IQ's appreciably below 50. For adequate data on IQ equivalents for the lowest IQ ranges, the results of retests from clinics and institutions for defectives are needed. Our present data are inadequate to afford any clinical evidence even of trends.

Unpublished data from a report submitted to the International Examination Inquiry Conference, Dinard, September, 1938, indicate that for a group of eight hundred seventy-four unselected Scottish children, including all children in Scotland born on any of four dates in 1926, the standard deviation of IQ's on the old scale is 15.58 points. The importance of this finding, based as it is on data secured from the only absolutely unselected group for which mental test results have

TABLE III.—IQ POINTS LOST OR GAINED ON REVISION RETEST COMPARED WITH IQ ON ORIGINAL STANFORD-BINET

L IQ	−25 or more	−20	−15	−10	−5	0	+5	+10	+15	+20	+25 or more	*N*
50 or below	1	5	1	4	1	....	1	....	...	...	...	13
55	3	...	4	2	2	1	1	...	...	...	...	13
60	3	3	5	20	1	7	1	...	...	...	...	40
65	7	7	7	11	13	7	1	...	...	...	...	53
70	7	11	13	23	15	29	4	3	3	...	...	108
75	7	13	14	23	27	14	9	5	3	...	...	115
80	12	18	12	35	30	50	13	11	1	...	1	183
85	6	11	28	29	42	24	25	15	2	2	...	183
90	...	12	25	36	34	27	13	26	7	2	...	182
95	1	3	20	27	41	29	19	20	8	2	1	171
100	2	2	9	23	26	27	25	15	16	3	4	152
105	1	1	...	12	23	19	25	15	6	5	1	108
110	...	...	2	5	12	14	10	14	3	9	5	74
115	...	...	...	2	4	13	7	8	5	5	4	48
120	...	...	...	....	1	4	4	8	2	1	1	21
125	...	...	...	1	1	1	4	5	6	2	2	22
130	...	...	...	....	....	....	1	4	4	1	6	16
135	...	...	...	....	....	....	....	1	...	4	1	6
140	...	...	...	....	....	....	1	...	2	1	1	5
145	...	...	...	....	....	....	....	...	...	...	1	1
150	...	...	...	....	....	....	....	...	...	...	1	1
155	...	...	...	....	....	....	....	...	...	...	1	1
>155	...	...	...	....	....	....	....	...	...	...	1	1
	50	86	140	253	273	266	164	149	68	37	31	1517
Per cent	3.3	5.7	9.2	16.7	18.0	17.5	10.8	9.8	4.5	2.4	2.1	100.0

ever been made available, makes it incumbent upon us to examine carefully the basis for other findings on variability of IQ scores. It is possible, as Terman has pointed out, that the smaller standard deviation for the old scale found for the original standardization group may be due to restricted range of the original population sampled. The population on whom the old scale was standardized was made up of California school children, whereas the revised scales are based on a much wider geographical and socio-economic sampling of cases. It is less easy to understand why the standard deviation of the IQ's on the

old scale differed in the same direction from the standard deviation on retests with the revised scales for the group of Kansas City school children. Possibly the testing itself was more carefully done with the new scale than with the old, more care taken in each case to determine a basal age and to make sure that the upper limit of successes had been thoroughly explored.

Basis for a comparison of the two scales at the upper age ranges is afforded by test results from the psychological clinic conducted by the writer at the Santa Clara County Juvenile Court. All cases who have been referred to the Juvenile Court during the past five years have been examined. IQ's for the first three hundred cases were based on the old scale; for the last two hundred on the revised scales. Although no retests are available, the two groups of cases are very similar in character, making valid comparisons possible between them. Obviously an adolescent group (the middle fifty per cent of our cases fall between ages fourteen and seventeen) should test higher on the revised scales than on the old scale for two reasons: (1) In computing $L - M$ IQ's for cases above thirteen an age correction has been introduced; and (2) the revised scales have more top, making it possible to attain higher mental age scores at the upper levels.

The two groups compared were not significantly different in age, socio-economic status, nationality, family disorganization, and a number of other factors, but did show a statistically significant difference between IQ's obtained on the old and new scales.

TABLE IV.—AGE COMPARISONS BETWEEN DELINQUENT GROUPS

	Delinquent No. 1	Delinquent No. 2
N	300	200
Mean	15.35	15.66
σ	2.15	2.27
σ_M	0.124	0.131
D/σ_D	1.7	

TABLE V.—X^2 CRITERION* OF THE SIGNIFICANCE OF DIFFERENCES IN BACKGROUND FACTORS BETWEEN DELINQUENT GROUPS

Occupation of fathers rated on Minnesota scale	$P = > .05$
Economic status rated on a three-point scale	$P = > .05$
Birthplace of parents: Foreign *vs.* American	$P = .02+$
Frequency of broken homes	$P = > .80$

* These differences are not statistically significant unless P is less than .02.

TABLE VI.—IQ COMPARISONS BETWEEN DELINQUENT GROUPS

	Delinquent No. 1	Delinquent No. 2
N	300	200
Mean	86.73	93.80
σ	18.56	19.70
σ_M	1.07	1.39
D/σ_D	4.0	

However, when the age correction is applied in computing the old scale IQ's of those in group I whose age is over thirteen and the significance of the difference between the corrected mean IQ's determined, the difference, while still in the same direction, is no longer found to be statistically significant ($D/\sigma_D = 2.02$). This indicates that only part of the obtained difference between mean IQ's is due to the age correction. The fact that grade placement, school achievement, and general information (as measured on the Judge Baker Foundation general information test) show no differences in achievement between the two groups would seem to indicate that the higher mean scores on the revised scales are due to the more accurate standardization of the new scales.

Because of the incomplete character of the clinical evidence so far available, it seems to the writer premature to suggest a descriptive classification for $L - M$ Stanford-Binet IQ's. That conviction is strengthened by past observation of the tyranny exerted by any printed classification system, however guardedly put forth, that enables us to put people into neat and convenient categories. Nevertheless, at the risk of falling into that very pitfall, let us turn to a consideration of some of the implications for such a classification of our fragmentary data on IQ trends.

The chief concern of the clinician is with the lower end of the scale. Whether one draws the borderline between *very superior intelligence* and *giftedness* at 140 or 150 is just a matter of deciding what percentage of the upper end of the distribution curve one wants to include. However, whether one calls an individual *feebleminded* or *dull normal* on the basis of an intelligence test does make an enormously greater difference than just deciding whether to draw the borderline at 60 or 70 on an IQ distribution curve. To make such a dividing line for classification purposes means to choose a point on the scale that serves

to delimit the group that is most likely to function as feebleminded from a group just beyond that point whose chances for normal functioning are greater.

As Terman pointed out in 1916 (4, p. 80):

> There are innumerable grades of mental deficiency ranging from somewhat below average intelligence to profound idiocy. In the literal sense every individual below the average is more or less mentally weak or feeble. Only a relatively small proportion of these, however, are technically known as feebleminded. It is therefore necessary to set forth the criteria as to what constitutes feeblemindedness in the commonly accepted sense of the word.

Sharp lines of distinction between normal intelligence and feeblemindedness do not correspond to clinical realities; there is no single point on the scale which can be designated as the dividing line between the legally and socially feebleminded on the one hand and the normal on the other. This fact has long been recognized by those who have been dealing with problems of mental deficiency. People of low intelligence, but not too low, may or may not be able to manage their affairs with ordinary prudence. It depends in part on many non-intellectual factors, such as environment, training, emotional balance, health, looks, economic conditions, etc.

Since it has been the practice of psychologists and others dealing with the mentally defective to classify as feebleminded, persons of low intelligence who, at maturity, are "incapable of successful personal-social adjustment without special assistance or supervision" (7, p. 334), and low intelligence alone has not been found a valid criterion of the ability to make social adjustments—get along in school, make an independent living, manage oneself and one's affairs—there has arisen much divergence of opinion concerning nomenclature and the determination of the borderlines of feeblemindedness. However, there is little or no disagreement concerning the fact that the term *feeblemindedness* implies not only low intelligence, but *usually* social inadequacy as well. Therefore, a clinically useful as well as psychologically valid schema for classification of IQ's must seek to differentiate measurable intellectual differences from the socio-psychological-legal distinctions that are important for the clinician who is dealing with practical problems of adjustment. To that end it is important to adopt for the lower end of the intelligence distribution a terminology that will serve to designate low intelligence without implying the presence of other behavior traits that have become associated with the term

feeblemindedness. There is some precedent for the adoption of *mental deficiency* as such a term.* In conformity with such a usage, a classification for the lower limits of intelligence need not depend on additional clinical evidence for its validity.

TABLE VII.—DISTRIBUTIONS OF COMPOSITE $L - M$ IQ'S OF THE STANDARDIZATION GROUP

IQ	*N*	Per cent	Classifications
160–169	1	0.03	Very superior
150–159	6	0.2	
140–149	32	1.1	
130–139	89	3.1	Superior
120–129	239	8.2	
110–119	524	18.1	High average
100–109	685	23.5	Normal or average
90– 99	667	23.0	
80– 89	422	14.5	Low average
70– 79	164	5.6	Borderline defective
60– 69	57	2.0	Mentally defective
50– 59	12	0.4	
40– 49	6	0.2	
30– 39	1	0.03	

Since classificatory terms are in common use and probably will continue to be applied to IQ categories, a classification which takes into account the above considerations and is based on the distribution of adjusted $L - M$ IQ's† for our standardization cases should prove

* The committee of experts who prepared the section on the handicapped child in the Report of the White House Conference on Child Health and Protection (7, pp. 329–390) suggests that a distinction be made between "mental deficiency" and "feeblemindedness," that "persons whose level of intelligence will never exceed a mental age of twelve years" be designated as "mentally defective," whereas in the determination of who is "feebleminded" mental defect and other social and psychological factors be taken into account.

† The composite $L - M$ IQ means were adjusted for the 1930 census frequencies of occupational groupings by dividing the subjects according to father's occupation into the seven categories of the Minnesota Scale for Occupational Classification, computing the mean IQ for each age level separately and then weighting according to the census frequencies (6, p. 36).

useful in orienting users of the revised scales. Such a distribution is presented in Table VII.

It will be noted that this classification makes no assumptions concerning the diagnostic significance of low IQ's for differentiating feeblemindedness. Of children with IQ's below 50, who in any case would be diagnosed as feebleminded and who, because of their low intelligence and consequent inability to profit by ordinary instruction, are not often found in the public schools, our standardization group includes only seven cases out of approximately three thousand examined in the schools. And for determining the developmental status of the children making up the group with IQ's between 50 and 69, usually designated as morons, we need more information. The chances that children with IQ's between 50 and 69 will function as feebleminded are greater than are their chances for normal functioning. For borderline defectives, as for the cases just below 70, it is especially important for mental diagnosis that the clinician have all the information usually made available through a complete case history as well as intelligence test data before attempting to differentiate such a psycho-social maladjustment as feeblemindedness.

REFERENCES

1. Anderson, Rose G.: "Methods and Results of Mental Surveys." *J. Appl. Psychol.*, 1922, Vol. VI, 1–28.
2. Bernreuter, Robert G. and Carr, Edward J.: "The Interpretation of IQ's on the *L* – *M* Stanford-Binet." *J. Educ. Psychol.*, 1938, Vol. XXIX, pp. 312–314.
3. Pintner, R.: *The Feebleminded Child.* Handbook of Child Psychology. Clark Univ. Press, 1938, pp. 802–841.
4. Terman, Lewis M.: *The Measurement of Intelligence.* Boston: Houghton Mifflin, 1916.
5. Terman, Lewis M., *et al.*: *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.* Baltimore: Warwick and York, 1917.
6. Terman, Lewis M. and Merrill, Maud A.: *Measuring Intelligence.* Boston: Houghton Mifflin, 1937.
7. White House Conference Report: *The Handicapped.* New York: Century, 1933.

A FACTOR ANALYSIS OF TEST-RETEST PERFORMANCE ON ATTITUDE AND ADJUSTMENT TESTS[1]

WALTER J. McNAMARA AND JOHN G. DARLEY

University of Minnesota

The problem of the constant or dynamic behavior of the individual in varying test situations underlies much of the work in the field of mental measurement. Intellectual capacity, specific aptitudes, and even broad occupational interest patterns are accepted as relatively constant characteristics of the individual; whereas, attitudes, adjustments, and learning as measured in achievement tests are assumed to be dynamic or changing aspects of behavior. The interpretations of the results of statistical analyses of test behavior tend to be confused not only by the traditional criteria of test construction—reliability and validity—but also by the unknown degree of constancy of the psychological behavior under analysis. The present paper makes no attempt to solve this basic problem; we shall merely present data which may be of interest to workers more familiar with theoretical implications.

These analyses have grown out of a broader study of the use of attitude and adjustment tests in a college personnel program.[2] During the academic year 1935–1936, three hundred twenty-six men and two hundred seventeen women in the General College of the University were tested and retested over an average interval of nine and two-tenths months, on the following attitude and adjustment scales: The Minnesota Scale for the Survey of Opinions,[3] the Bell Adjustment Inventory,[4] and the Minnesota Inventories of Social Attitudes.[5] These schedules yield thirteen attitude or adjustment scores. The

[1] The authors are indebted to Dr. John E. Anderson and Mr. Russell C. Smart for their assistance and critical review of the data in this study, and to Miss Blanche Anderson for statistical help.

[2] Darley, John G.: *An Analysis of Attitude and Adjustment Tests, with Special Reference to Conditions of Change in Attitudes and Adjustments.* Ph. D. Thesis on file at the University of Minnesota Library.

[3] Rundquist, E. A. and Sletto, R. F.: *Scoring Instructions for the Minnesota Scale for the Survey of Opinions.* University of Minnesota Press, 1936.

[4] Bell, H. M.: *Manual for the Adjustment Inventory.* Stanford University Press, 1934.

[5] Williamson, E. G. and Darley, J. G.: *Manual for the Minnesota Inventories of Social Attitudes.* The Psychological Corporation, New York City, 1937.

Minnesota Scale for the Survey of Opinions is scored for: Morale; inferiority feelings; family adjustments; attitudes toward the law; economic conservatism; attitudes toward education; and general adjustment, an intra-test scale of items differentially weighted for a measure of general satisfactoriness of individual adjustments. The Bell Adjustment Inventory is scored for home, health, social, and emotional adjustments. The Minnesota Inventories of Social Attitudes yield separate scores for social preference and social behavior. In the total groups are ninety-seven men and fifty-five women who had left the University or transferred to another college within the University during the year 1935–1936, and who coöperated in a follow-up study carried on in the Fall quarter of the year 1936–1937.

To provide samples for more intensive study, one hundred men and one hundred women were selected at random from the larger samples of three hundred twenty-six men and two hundred seventeen women. The means and standard deviations of the smaller samples on the original testing showed no significant variation from the same statistics for the larger samples from which they were drawn. As a further check on the comparability of the small and large samples, it was found that the percentage of follow-up cases in the small sample corresponded quite closely with the same percentage in the larger sample. Since in the larger study the follow-up cases had shown some differentials in both test and retest performance, as compared to the group remaining in school, it was advisable to determine that the small sample was not by chance overloaded with follow-up cases. These two checks seemed to afford satisfactory evidence that the samples of one hundred men and one hundred women were typical of the larger populations from which they were drawn.

Prior to carrying out a factor analysis of the intercorrelations for each sex on the original testing and again on the retesting, the amount of individual and group change over the average interval of nine and two-tenths months was studied. Table I summarizes these data. Odd-even reliability coefficients, corrected by the Spearman-Brown formula, were determined for both the test and retest situations. These reliabilities correspond quite closely to the reliabilities obtained by the authors of the various tests in the process of standardization. Test-retest correlations were calculated as measures of individual changes in relative performance from test to retest. Finally, critical ratios between test and retest means were calculated to show average changes on the various tests.

Test-retest correlations for the larger samples of three hundred twenty-six men and two hundred seventeen women (not shown here) were closely similar to the same correlations appearing in Table I. The critical ratios for the larger samples are somewhat different. Setting a level of 2.5 for these ratios, three hundred twenty-six men yielded eight critical ratios above this level, and two hundred seventeen women yielded ten such ratios. As shown in Table I, the subsamples of one hundred gave one critical ratio above 2.5 for men and eight critical ratios above 2.5 for women. In each case, however, the *absolute* mean differences for the large and small samples are of approximately the same magnitude. As is obvious, the small sampling for one hundred men involved the selection of about one out of three cases, and for one hundred women one out of two cases. Thus the critical ratios for men are more affected by the decrease in number of cases than are the critical ratios for women.

TABLE I.—CORRECTED ODD-EVEN RELIABILITY OF TEST AND RETEST SITUATIONS, TEST-RETEST CORRELATIONS AND TEST-RETEST CRITICAL RATIOS FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS

Test	One-hundred men				One-hundred women			
	Corrected test reliability	Corrected retest reliability	Test-retest correlation	Test-retest critical[1] ratio	Corrected test reliability	Corrected retest reliability	Test-retest correlation	Test-retest critical[1] ratio
Minnesota scale for survey of opinions.								
Morale	.744	.868	.647	2.36	.810	.827	.634	4.42
Inferiority	.810	.866	.609	4.14	.886	.868	.526	4.83
Family	.811	.899	.639	1.24	.887	.897	.761	2.94
Law	.768	.775	.554	1.14	.807	.849	.574	2.88
Economic conservatism	.868	.866	.792	0.66	.819	.849	.590	0.44
Education	.662	.834	.462	1.77	.830	.856	.628	3.33
General adjustment	.774	.778	.609	1.38	.814	.786	.639	2.58
Bell adjustment inventory.								
Home	.790	.851	.713	−0.62	.888	.902	.816	0.66
Health	.716	.733	.716	0.44	.744	.797	.810	−0.89
Social	.905	.905	.838	0.32	.885	.901	.778	2.97
Emotional	.797	.819	.681	0.93	.853	.849	.695	2.30
Minnesota inventories of social attitudes.								
Social preferences	.911	.926	.727	0.95	.894	.909	.620	2.16
Social behavior	.904	.926	.645	1.39	.871	.843	.689	4.06

[1] Based on *standard* errors of mean differences, and calculated from formula for standard error of difference in correlated data. A minus sign indicates change in an unfavorable direction from test to retest

Table I shows also a differential rate of change among the tests as well as from one sex to another. The effect of changing reliability of the tests is difficult to determine. Changing reliability does not account for all observed changes, however, since the social preference and social behavior tests show a clear sex difference accompanying relatively high and stable reliabilities. The sex difference may be shown also in the Spearman rank-order correlation between men's and women's test-retest coefficients. This coefficient is .394, indicating considerable divergence in the relative ranks of the thirteen tests as far as test-retest stability is concerned. Men, for example, are much more constant over a period of time in their economic attitudes than are women. The possible interpretation of these changes has been discussed elsewhere,[1] and is not a primary concern at this time. We are trying to establish the presence of the change itself, among the tests and between the sexes, in order to observe the effect of the change in the results of the factor analysis. The application of either the correction for attenuation or Paulsen's infrequently-used Coefficient of Trait Variability[2] to the test-retest correlations in Table I results in few relative changes in order of magnitude of the correlations; the interpretation of the corrected coefficients in respect to areas of change is no clearer than the interpretation of the uncorrected coefficients.

Tables II, III, IV and V present the correlational matrices for men on test and retest and women on test and retest, respectively. Tables VI, VII, VIII and IX present the results of the factor analysis of the matrices using Thurstone's centroid method of factor analysis as described and illustrated by Guilford.[3]

In the analysis of Table II the communalities of the tests were estimated, using the highest correlation in the column for that test. Then a second analysis of the data was made using as the communalities the values computed from the first analysis, but the resultant factor loadings were so similar that only a single analysis of Tables III, IV, and V was made, using the highest r in each column as the communality. The factor loadings presented in Tables VI, VII, VIII,

---

[1] Darley, J. G.: "Changes in Measured Attitudes and Adjustment." *Journal of Social Psychology*, to appear in a forthcoming issue.

[2] Paulsen, G. B.: "The Reliability and Consistency of Individual Differences in Motor Control." *Journal of Applied Psychology*, Vol. XIX, No. 1, February, 1935, pp. 29–42, and Vol. XIX, No. 2, April, 1935, pp. 166–179.

[3] Guilford, J. P.: *Psychometric Methods*. McGraw-Hill Book Company, 1936, Chapter XIV.

and IX were obtained by rotating the axes in one plane at a time until reference axes were obtained that seemed to give the most psychological meaning.

TABLE II.—INTERCORRELATIONS FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED MEN—ORIGINAL TEST

	Inferiority	Family	Law	Economic conservatism	Education	General adjustment	Home	Health	Social	Emotional	Social preferences	Social behavior
Morale	.50	.30	.48	.19	.43	.74	.26	.08	.29	.34	.20	.30
Inferiority	...	.17	.24	.12	.23	.45	.17	−.06	.52	.28	.22	.48
Family	...	...	.37	−.16	.32	.40	.40	.22	.14	.09	.29	.18
Law	...	...	...	.21	.40	.58	.15	.06	.13	.28	.06	.14
Economic conservatism	...	...	...	.....	.08	.39	.06	.02	.17	.22	.20	.14
Education	...	...	...	.....	...	.52	.10	−.06	−.15	.01	−.06	−.04
General adjustment	...	...	...	.....	...	...	.25	.11	.28	.30	.15	.29
Home	...	...	...	.....	...	...	...	.32	.31	.48	.32	.18
Health	...	...	...	.....	...	...	...	.....	.16	.38	.20	.13
Social	...	...	...	.....	...	...	...	.....	.....	.46	.43	.69
Emotional	...	...	...	.....	...	...	...	.....	.....	...	.20	.32
Social preferences	...	...	...	.....	...	...	...	.....	.....	...	.....	.49

It will be observed in the tables of intercorrelations that some of the intercorrelations change, for both sexes, from the test to the retest situation. These changes are slightly greater in the coefficients of lower numerical value, but result in little change in psychological interpretations. However, in the factor loadings, these numerical changes seem to produce marked differences in interpretation from test to retest. From Table I we may infer that the factor loadings for women will be less stable than for men, since women show greater changes from test to retest. But it seems that the factor loadings of the tests should not change quite so much as they do in the nine-month period if they are entities of basic psychological meaning. If a factor may be conceived psychologically as a constellation of behavior tendencies of stability in individuals, differing from group to group only in strength (or amount), but occupying the same relative position

in the individuals' total behavior repertory, we should not expect the mathematical operations of factor analysis to result in the diminution or augmentation of a constellation of behavior tendencies (or a factor so defined) by excluding or including certain tests in the retest situation.

TABLE III.—INTERCORRELATIONS FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED MEN—RETEST

	Inferiority	Family	Law	Economic conservatism	Education	General adjustment	Home	Health	Social	Emotional	Social preferences	Social behavior
Morale	.63	.42	.54	.14	.47	.76	.35	−.01	.34	.31	.31	.34
Inferiority	...	.32	.30	.18	.23	.57	.35	.14	.60	.55	.38	.61
Family	...	...	.38	−.03	.24	.38	.60	.04	.07	.22	.25	.18
Law	...	...	...	.29	.44	.55	.24	.08	.05	.26	.12	.14
Economic conservatism	...	...	...	.....	.05	.40	.12	−.09	.04	.08	.11	.06
Education	...	...	...	.....	...	.54	−.02	−.15	.11	−.04	.13	.16
General adjustment	...	...	...	.....	...	...	.25	−.09	.34	.19	.22	.29
Home	...	...	...	.....	...	...	.....	.28	.26	.58	.18	.22
Health	...	...	...	.....	...	...	.....	.....	.12	.41	.03	.16
Social	...	...	...	.....	...	...	.....	.....	...	.54	.49	.77
Emotional	...	...	...	.....	...	...	.....	.....	...	.....	.23	.43
Social preferences	...	...	...	.....	...	...	.....	.....	...	.....	...	.66

If, on the other hand, a factor represents a constellation of behavior tendencies that may be sufficiently dynamic to be markedly influenced by stimuli incident to the passage of time in the individuals' lives, then the mathematical operations of factor analysis are simply reflecting psychological phenomena as yet inadequately analyzed. In which of these categories fall the behavior tendencies measured by attitude and adjustment tests of this nature? A third possibility may also be considered here. It may be that, while these tests were identical at the beginning and end of the school year, the personalities of the individuals taking them had changed to such an extent that the tests came to measure not only dynamic, but *structurally different*, behavior tendencies. If a test of induction on one age level may become a test of perception on another age level, it is conceivable that a non-specific,

impersonal, attitude toward law may become, by unpleasant contact with the legal system, an emotionally-charged measure of personal maladjustment at another time. Thus retests of personality characteristics may be more like new tests for the individual than like the original tests.[1]

TABLE IV.—INTERCORRELATIONS FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED WOMEN—ORIGINAL TEST

	Inferiority	Family	Law	Economic conservatism	Education	General adjustment	Home	Health	Social	Emotional	Social preferences	Social behavior
Morale	.58	.54	.69	.38	.56	.86	.36	.04	.25	.32	.20	.17
Inferiority	...	.40	.41	.28	.36	.50	.32	.15	.47	.42	.29	.50
Family	...	...	.50	.19	.32	.60	.58	.11	.23	.21	.02	.15
Law	...	...	...	.49	.51	.67	.27	.10	.09	.18	.16	.02
Economic conservatism	...	...	...	...	.18	.38	.14	.07	.08	.04	.19	.18
Education	...	...	...	...	...	.58	−.02	−.08	.07	.11	.04	.01
General adjustment	...	...	...	...	...	...	.33	.05	.19	.28	.13	.15
Home	...	...	...	...	...	...	.....	.38	.23	.50	.11	.20
Health	...	...	...	...	...	...	.....	.....	.05	.38	.06	.06
Social	...	...	...	...	...	...	.....	.....	...	.42	.42	.67
Emotional	...	...	...	...	...	...	.....	.....	...	...	.15	.34
Social preferences	...	...	...	...	...	...	.....	.....	...	...	...	.58

Still a fourth alternative must be considered. How many cases are necessary to provide a correlational matrix of sufficient stability to rule out the effect of chance correlational errors on the operations of a factor analysis? In this study samples of one hundred have been used. The intercorrelations for these samples change from test to retest, as may be seen. These changes have little effect on the psychological interpretations of the correlations themselves, but may be a marked influence on the data obtained from factor analysis. If this is true, the psychological phenomenon may not warrant such a rigorous test as factor analysis, or it may be that factor analysis should

[1] Both Dr. Thurstone and Mr. Smart, in independent conversations, suggested this possible interpretation to the authors.

be applied only to data drawn from samples so extensive as to be thoroughly typical of defined parameters.

TABLE V.—INTERCORRELATIONS FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED WOMEN—RETEST

	Inferiority	Family	Law	Economic conservatism	Education	General adjustment	Home	Health	Social	Emotional	Social preferences	Social behavior
Morale	.64	.47	.62	.12	.56	.80	.26	.10	.14	.33	.30	.34
Inferiority	...	.45	.33	.02	.19	.52	.33	.14	.50	.50	.29	.62
Family	...	...	.37	.23	.19	.54	.71	.14	.20	.43	.20	.25
Law	...	...	...	.27	.42	.62	.16	−.02	−.03	.14	.24	.15
Economic conservatism	...	...	...	...	−.07	.36	.04	−.26	−.01	−.10	.18	.11
Education	...	...	...	...	.....	.54	.06	−.02	−.14	.14	.05	.04
General adjustment	...	...	...	...	.....	...	.33	−.03	.08	.33	.14	.32
Home	...	...	...	...	.....	...	...	.39	.26	.65	.00	.23
Health	...	...	...	...	.....	...	...	.....	.17	.53	.06	.14
Social	...	...	...	...	.....	...	...	.....	.....	.43	.28	.72
Emotional	...	...	...	...	.....	...	...	.....	.....	.....	.05	.40
Social preferences	...	...	...	...	.....	...	...	.....	.....	.....	...	.40

We may illustrate the foregoing points more specifically by reference to the test names that tend to group together as a result of the factor analysis. The decision as to the magnitude of a factor loading that warrants its test's inclusion in a psychologically meaningful context must in the last analysis be largely determined by one's familiarity with the tests themselves. The present groupings, then, are based in part on clinical experience with the tests. In addition to this clinical judgment, Thurstone's measure of relative importance ($\Sigma K^2/N$) has been calculated for each factor and has been used as a guide in listing each factor in columnar order in Tables VI, VII, VIII, and IX.

In the case of men, the following four tests are grouped in a meaningful context in both test and retest situations: General adjustment, morale, education, law. The items of the general adjustment test are so heavily drawn from the items of the morale test as to represent duplicated measures. This constellation, or first factor, may be

named a generalized respect for and adjustment to authority as imposed by society on the individual. The second relatively important factor for men in the original testing is seen in the loadings for the inferiority, social adjustment, and social behavior tests; at the time of retesting, this factor is still second in relative importance, but its composition has shifted slightly, to include more of the test of social preferences and less of the test of inferiority feelings. This general entity appearing at both testing and retesting may be labelled as socialized interests and socially skillful or adequate behavior.

TABLE VI.—FACTOR LOADINGS, COMMUNALITIES, RELIABILITIES, UNIQUENESS, AND SPECIFICITY FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED MEN—ORIGINAL TEST

Test	I	II	III	IV	V	h^2	Reliability	Uniqueness	Specificity	Unreliability
Minnesota scale for survey of opinions.										
Morale	.680	.379	.028	.187	.115	.655	.744	.345	.089	.256
Inferiority	.338	.680	.012	.011	.027	.578	.810	.422	.232	.190
Family	.450	.032	.689	−.032	−.118	.693	.811	.307	.118	.189
Law	.618	.100	.069	.148	.126	.434	.768	.566	.334	.232
Economic conservatism	.129	.059	−.120	.145	.624	.445	.868	.555	.423	.132
Education	.700	−.089	.027	−.051	−.003	.501	.662	.499	.161	.338
General adjustment	.814	.274	−.014	.216	.332	.895	.774	.105	−.121	.226
Bell adjustment inventory.										
Home	.151	.131	.467	.454	.005	.464	.790	.536	.326	.210
Health	−.033	−.017	.334	.460	.009	.325	.716	.675	.391	.284
Social	−.065	.813	.232	.210	.164	.790	.905	.210	.115	.095
Emotional	.107	.315	.154	.745	.070	.694	.797	.306	.103	.203
Minnesota inventories of social attitudes.										
Social preferences	−.044	.327	.514	.032	.341	.490	.911	.510	.421	.089
Social behavior	.008	.695	.307	.017	.256	.643	.904	.357	.261	.096
$\Sigma K^2/N$	.182	.159	.097	.088	.058					

The third important factor in the original tests for men is difficult to label. It comprises the family test, social preferences, and to some extent the home adjustment test. By the time of retesting, this factor, which is fourth rather than third in importance, shows a more homogeneous composition, with the family and home adjustment scales showing the greatest saturation. Whatever entity is being measured seems to relate primarily to familial adjustment, and is psychologically meaningful in late adolescence, when the individual

may be in process of establishing at least a psychic independence in the family group.

The fourth factor in the original testing appears most clearly in the emotional adjustment test, and to a lesser extent in the health and home adjustment measures. In the retesting, not only the emotional adjustment test, but also the social adjustment, the inferiority, and the health adjustment tests show a saturation of this factor, which is now third, rather than fourth, in relative importance. These groupings might comprise evidence for the more familiarly known neurotic tendencies; the changes in saturation may indicate that neurotic tendencies, or emotionality, are part of a larger context in which other and more specific adjustments are contributing phenomena.

TABLE VII.—FACTOR LOADINGS, COMMUNALITIES, RELIABILITIES, UNIQUENESS, AND SPECIFICITY FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED MEN—RETEST

Test	I	II	III	IV	V	h^2	Reliability	Uniqueness	Specificity	Unreliability
Minnesota scale for survey of opinions.										
Morale	.727	.269	.202	.134	.262	.728	.868	.272	.140	.132
Inferiority	.380	.444	.481	.150	.230	.648	.866	.352	.218	.134
Family	.389	.126	−.062	.644	.028	.587	.899	.413	.312	.101
Law	.629	.035	.040	.233	.253	.517	.775	.483	.258	.225
Economic conservatism	−.034	.063	−.108	.012	.791	.643	.866	.357	.223	.134
Education	.682	.202	−.074	−.103	−.060	.526	.834	.474	.308	.166
General adjustment	.765	.262	.049	.025	.441	.851	.778	.149	−.073	.222
Bell adjustment inventory.										
Home	.133	.055	.262	.762	.150	.693	.851	.307	.158	.149
Health	−.121	−.086	.476	.256	−.009	.314	.733	.686	.419	.269
Social	.047	.636	.543	.004	.143	.722	.905	.278	.183	.095
Emotional	.047	.095	.739	.361	.244	.747	.819	.253	.072	.181
Minnesota inventories of social attitudes.										
Social preferences	.064	.742	.051	.220	.021	.606	.926	.394	.320	.074
Social behavior	.062	.840	.378	.088	.047	.862	.926	.138	.064	.074
$\Sigma K^2/N$	.178	.160	.121	.104	.086					

The fifth factor, in both test and retest, shows a heavy and specific concentration in the lone measure of economic conservatism. It is amazing only that this factor should be so specific within one test, but this apparent specificity may be due to the absence of similar additional measures in the battery.

In the case of men, then, the factor loadings are fairly consistent from the test to retest situations, with the indicated changes in relative importance. Men also, in Table I, show less mean change on the attitude and adjustment tests than do women, although the test-retest correlations for both sexes are of about the same magnitude. The factor analyses for women are considered next.

TABLE VIII.—FACTOR LOADINGS, COMMUNALITIES, RELIABILITIES, UNIQUENESS, AND SPECIFICITY FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED WOMEN—ORIGINAL TEST

Test	I	II	III	IV	V	h^2	Reliability	Uniqueness	Specificity	Unreliability
Minnesota scale for survey of opinions.										
Morale	.870	.119	.038	.156	.226	.848	.810	.152	−.038	.190
Inferiority	.520	.486	.056	.238	.113	.579	.886	.421	.307	.114
Family	.586	.076	.606	−.056	.112	.732	.887	.268	.155	.113
Law	.672	−.082	.019	.160	.520	.755	.807	.245	.052	.193
Economic conservatism	.281	.094	.034	−.027	.557	.400	.819	.600	.419	.181
Education	.710	−.013	−.114	.076	.027	.524	.830	.476	.306	.170
General adjustment	.878	.040	.100	.128	.187	.834	.814	.166	−.020	.186
Bell adjustment inventory.										
Home	.203	.164	.678	.386	.110	.689	.888	.311	.199	.112
Health	−.094	.045	.249	.534	.145	.379	.744	.621	.365	.256
Social	.199	.786	.068	.120	−.112	.689	.885	.311	.196	.115
Emotional	.205	.335	.220	.572	−.042	.532	.853	.468	.321	.147
Minnesota inventories of social attitudes.										
Social preferences	.010	.593	−.095	.035	.373	.501	.894	.499	.393	.106
Social behavior	.066	.867	.078	−.023	.112	.775	.871	.225	.096	.129
$\Sigma K^2/N$	.255	.164	.076	.070	.069					

The following tests, in the original testing for women, show a heavy loading of a first factor: Morale; inferiority; family; law; education; general adjustment. At the time of retesting, the tests of morale, law, education, and general adjustment and possibly inferiority are loaded most heavily with this factor. May this factor be considered the same as the first factor in the case of men: A generalized (and more pervasive) respect for and adjustment to authority as imposed by society on the individual?

The second factor of relative importance appears most significantly in the original tests of social preference, social behavior, social adjustment, and inferiority. With the exception of social preferences, the

same tests recur with heavy loadings of this factor in the retesting, but rank third rather than second in relative importance. This constellation of behavior tendencies is similar to the second factor loadings described for the men and may refer to the same type of social behavior. The third factor in the original tests for women shows the heaviest loading in the family and home adjustment tests. But in the retesting, not only the family and home adjustment scales, but also the general adjustment scale and the economic conservatism scale have become part of this same psychological context; in addition, the factor moves from third to fourth place in relative importance. One is at a loss to find a generalized rubric for these four tests.

TABLE IX.—FACTOR LOADINGS, COMMUNALITIES, RELIABILITIES, UNIQUENESS AND SPECIFICITY FOR THE THIRTEEN ATTITUDE AND ADJUSTMENT TESTS—ONE HUNDRED WOMEN—RETEST

Test	I	II	III	IV	V	h^2	Reliability	Uniqueness	Specificity	Unreliability
Minnesota scale for survey of opinions.										
Morale	.830	.126	.136	.222	.346	.892	.827	.108	−.065	.173
Inferiority	.433	.336	.500	.151	.218	.621	.868	.379	.247	.132
Family	.232	.416	.040	.720	.048	.749	.897	.251	.148	.103
Law	.571	−.081	.003	.381	.308	.573	.849	.427	.276	.151
Economic conservatism	−.009	−.121	.103	.645	−.002	.441	.849	.559	.408	.151
Education	.666	.021	−.105	.035	.036	.458	.856	.542	.398	.144
General adjustment	.798	.057	.165	.500	−.031	.918	.786	.082	−.132	.214
Bell adjustment inventory.										
Home	.042	.713	.017	.498	−.039	.760	.902	.240	.142	.098
Health	−.035	.621	−.058	−.111	.229	.455	.797	.545	.342	.203
Social	−.069	.358	.744	.032	.105	.699	.901	.301	.202	.099
Emotional	.227	.804	.204	.098	.026	.750	.849	.250	.099	.151
Minnesota inventories of social attitudes.										
Social preferences	.067	.061	.312	.161	.528	.410	.909	.590	.499	.091
Social behavior	.141	.230	.854	.116	.170	.844	.843	.156	−.001	.157
$\Sigma K^2/N$	.186	.158	.134	.132	.049					

The emotional scale and the health scale are heavily loaded with a fourth factor in the original testing, and are meaningful in this context. But in the retesting, four tests fall together: Emotional, health, home, and family. Furthermore, the factor moves from fourth to second place in relative importance. Has the constellation of behavior tendencies included in the emotional and health tests spread to include

the home and family tests? Or are the second and fourth factors in the retesting "crossed" in some manner, with the home and family scales as a basic connecting link for two separate constellations of behavior tendencies?

The fifth factor in the original testing is found in the economic conservatism and law scales, and may be a general tendency to conservatism or radicalism. But there is no comparable factor in the retesting; instead, the test of social preferences emerges as a comparatively lonely residual factor.

Admittedly, some of the behavior tendencies in women that are tapped by these attitude and adjustment tests are less stable than they are for men. Evidence for this may be found in the data of Table I and in the tables of intercorrelations. The factor analyses of the intercorrelations for women reflect this lack of constancy. However, if we assume that factor analysis will slowly isolate for more intensive study the myriad constellations of behavior tendencies that characterize the human organism, it would seem that this assumption needs to be qualified or further defined in terms of the degree of constancy that may be found in various areas of behavior. It would be difficult, for example, to construct new tests for women on the basis of factor analysis of the original tests, without taking into consideration the kaleidoscopic changes in groupings of tests from the test to retest situation.

Many other research possibilities arise from these data. The authors plan to group the present scales and rescore them for a new item analysis that would represent a greater homogeneity of items for further use. Other forms of factor analysis might be applied to the tables of intercorrelations. It is hoped that the data presented herewith may be used again in the study of the basic problem of changes in measured behavior as they affect the theory and use of factor analysis methods.

THE EFFECT OF TEACHER POPULARITY UPON ATTITUDE TOWARD SCHOOL SUBJECTS

STEPHEN M. COREY AND GEORGE S. BEERY

University of Wisconsin

INTRODUCTION

College administrators are apparently rather generally agreed that one of the less valuable criteria for the evaluation of teaching effectiveness is the judgment of students.[11,p.24] High-school principals and superintendents do not hold consumer opinion in such low regard. Hart,[8] in his collection of secondary-school pupils' judgments of their teachers, is explicit in his emphasis upon the importance of pupil regard for their instructors. Recently a number of papers have been reported dealing with pupil ratings of teachers[2,6,10,12,13] which imply the importance of such judgments.

To the writers, student opinion of the effectiveness of teaching is of importance largely because it reflects teacher-pupil relationships. While it is true that occasionally an adult will report that a high-school or college teacher who at the time was disliked, actually accomplished much as an instructor, such judgments are the exception. In all likelihood they are remembered for that very reason.

Personal popularity on the part of the teacher can be striven for in a number of ways. Those teachers who do not attain a satisfactory degree of such popularity are quite apt to malign their more favored colleagues by frequent references to "vaudeville" tactics and lowered standards. That a large majority of the very popular high-school teachers do not degrade themselves by such practices seems fairly well established. While there is at least some possibility of a "halo" operating, there are surprisingly few trait differences between teachers pupils like best and teachers pupils admit have taught them most.[8]

It is conceivable that personal popularity with high-school pupils might be one valid criterion of teaching effectiveness. Liking a teacher might be closely related to learning. So far as the writers know, there have been no investigations of the relationship between teacher popularity and teaching effectiveness as judged either by supervisors, or more validly, by measured pupil progress. The other factors which were once thought to be essential for teaching success seem not to have justified our confidence.[1]

METHODS AND SUBJECTS

Because of their respect for student ratings of teacher effectiveness, the writers have attempted to determine the relationship between attitude toward a subject and liking for the teacher giving instruction in that subject. For this purpose, two different inquiries were made of one hundred fifty-two university students at the beginning of their freshman year—a time when recollections of high-school experiences would be clearest. In September these freshmen were requested to:

1. List by name the teacher or teachers who were liked best during high school. If no teacher was outstandingly popular none need be mentioned.
2. Indicate after each of these teachers the subject or subjects he taught.
3. List by name the teacher or teachers who were liked least during high school. If no teacher was outstandingly unpopular none need be mentioned.
4. Indicate after each of these teachers the subject or subjects he taught.

One month later, after sufficient time had lapsed since the first inquiry to cause most of the freshmen to forget it entirely, the students were asked to:

1. List the high-school subjects that were liked best.
2. Check those liked subjects which are being taken now or which will be taken in the university, and indicate whether they are or will be studied because they are required, only, or because of choice.
3. List the high-school subjects liked least.
4. Check those disliked subjects that are taken or will be taken in the university and indicate whether this is because of choice or because of requirements.

This second inquiry did not involve the mention of any teachers' names. Inasmuch as both questionnaires were signed, however, it was possible to determine the instructors in the best liked and least liked subjects by referring to the September report. No attempt was made to assign any precise meaning to the word "like." It was employed in its popular sense and the forthright manner in which the students reacted indicated no confusion on their parts. Nor did there appear to be any particular reluctance to indicate by name on a signed paper the teachers who were disliked.

The responses on the two papers were tabulated and summarized and comparisons made to provide answers for these questions:

1. How many high-school subjects were extremely well liked?
2. How many high-school subjects were greatly disliked?

3. How many high-school teachers were extremely well liked?
4. How many high-school teachers were greatly disliked?
5. How many times were both teacher and the subject he taught extremely well liked?
6. How many times was the teacher well liked and the subject he taught disliked?
7. How many times was the subject liked and the person teaching the subject disliked?
8. How many times were both teacher and the subject he taught disliked?
9. What effect did liking or disliking the teacher or subject have upon continued study of the subject in the university?

RESULTS

Table I summarizes the student replies to the two inquiries. The results from the two inquiries do not correspond in every particular because neither all subjects nor all teachers were particularly well liked or particularly disliked.

It is apparent from Table I that there were more unpopular subjects than there were unpopular teachers. Out of seven hundred two teachers mentioned by name, four hundred ninety or seventy per cent were extremely well liked. Of the nine hundred fifty-four times that school subjects were named as being liked or disliked, five hundred sixty-nine, or sixty per cent, were placed in the first category. This, in itself, is worthy of comment even though it must be borne in mind that there might have been some tendency to refuse to name disliked teachers. On the average, each freshman reported that approximately three of his high-school teachers were very popular as compared to approximately one who was heartily disliked.

There was a very pronounced tendency for the freshmen to report a strong liking for both teacher and subject if either was particularly popular. There were only eighty-two cases where both teacher and subject were disliked as compared with two hundred ninety-two cases where both were popular. There were twice as many popular teachers teaching unpopular subjects as there were popular subjects taught by unpopular teachers. Apparently high-school pupils can still enjoy a teacher even though his subject is unpopular. If the teacher is not liked, however, the attitude is very apt to carry over to the subject. Each of these tendencies would seem to support Dewey's claim:[3,p.59] "With the young, the influence of the teacher's personality is intimately fused with that of the subject; the child does not separate or even distinguish the two."

TABLE I.—THE RELATIONSHIP BETWEEN THE POPULARITY OF HIGH-SCHOOL TEACHERS AND THE POPULARITY OF THE HIGH-SCHOOL SUBJECTS THEY TEACH AS REPORTED BY ONE HUNDRED FIFTY-TWO UNIVERSITY FRESHMEN

1. Total number of teachers well liked		490
Average number per student		3.2
2. Total number of teachers greatly disliked		212
Average number per student		1.4
3. Total number of subjects well liked		569
Average number per student		3.7
4. Total number of subjects greatly disliked		385
Average number per student		2.5
5. Total number of times both the teacher and his subject were well liked		292
a. Total number of times students planned to continue in university with well liked subject taught by well liked teacher		198
Because of student's choice	190	
Because of course requirements	8	
6. Total number of times the teacher was liked but the subject disliked		65
a. Total number of times students planned to continue in university with disliked subject taught by well liked teacher		17
Because of student's choice	3	
Because of course requirements	14	
7. Total number of times the teacher was disliked but the subject liked		31
a. Total number of times students planned to continue in university with popular subject taught by unpopular teacher		8
Because of student's choice	4	
Because of course requirements	4	
8. Total number of times both the teacher and the subject he taught were disliked		82
a. Total number of times students planned to continue in university with disliked subject taught by disliked teacher		22
Because of student's choice	4	
Because of course requirements	18	

With respect to the relationship between liking the teacher or subject and continuation with the subject in college the data in Table I make possible some very interesting inferences. College students, speaking generally, pursue certain studies for either one of two reasons; namely, because the subject is required for a certain curricular sequence or because it is elected. It seems that disliking a high-school teacher,

even though the subject he taught was liked, is enough to discourage pupils from electing to continue with the study of that subject in college. Out of thirty-one instances of popular high-school subjects taught by unpopular teachers there were only four cases (thirteen per cent) where the subject was *elected* for continued study in college. On the other hand, where both teacher and subject were well liked (two hundred ninety-two cases) there were one hundred ninety instances (sixty-five per cent) of electing to continue with the subject in college. When neither the subject or teacher was liked (eighty-two cases) only four students (five per cent) elected to continue the subject in college.

DISCUSSION

It is, of course, impossible to state definitely whether disliking the teacher leads to disliking the subject or disliking the subject leads to dislike of the teacher. Unpopularity on the part of either one would very probably carry over and effect the other. The same would be true, of course, in the case of popularity. The teacher of the well-liked subject has the dice loaded in his favor, and certainly the popular teacher frequently, if not usually, engenders liking for his subject.

We are not altogether in the dark as regards the relationship between liking a subject and succeeding as a student of it. Common-sense evidence insists that the two are related and there are some objective studies supporting the same point of view. Fryer[7] and others[4,5,9,14] have reported rather substantial positive correlations between estimated "interest" in a school subject and achievement as measured by academic grades.

If, as our data imply, liking for a subject is rather closely related to liking the teacher, and bearing in mind that interest in a subject is, as indicated above, one criterion for predicting achievement, we educators are a bit remiss when we appear to be but occasionally concerned with the personal popularity of teachers with pupils. "Getting along" with the boys and girls has long been recognized as being important, but chiefly from the point of view of maintaining discipline. The teacher who maintains discipline, however, may be quite a different creature from the teacher of whom the children are fond. Popular teachers have few disciplinary cases, but it does not follow that absence of disciplinary disturbances implies teacher popularity—or pupil learning either for that matter.

BIBLIOGRAPHY

1. Corey, Stephen M.: "What are the factors involved in the success of high-school teachers?" *North Central Association Quarterly*, Vol. x, 1935, pp. 224–231.
2. Davies, J. E.: "What are the traits of the good teacher from the standpoint of junior-high-school pupils?" *School and Society*, Vol. xxxviii, 1933, pp. 649–652.
3. Dewey, John: *How We Think* (Revised Edition). New York: D. C. Heath and Company, 1933, 301 pp.
4. Dunlap, Jack W.: "Preferences as indicators of specific academic achievement." *Journal of Educational Psychology*, Vol. xxvi, 1935, pp. 411–415.
5. ———: "The predictive value of interest test items for achievement in various school subjects." *Journal of Applied Psychology*, Vol. xix, 1935, pp. 53–58.
6. Flory, C. D.: "Personality rating of prospective teachers." *Educational Administration and Supervision*, Vol. xvi, 1930, pp. 135–143.
7. Fryer, Douglas: "Interest and ability in educational guidance." *Journal of Educational Research*, Vol. xvi, 1927, pp. 27–39.
8. Hart, Frank W.: *Teachers and Teaching.* New York: The Macmillan Company, 1934, 285 pp.
9. King, Irving and Adelstein, Morris: "The permanence of interests and their relation to abilities." *School and Society*, Vol. vi, 1917, pp. 359f.
10. Knoble, J. W.: "What the student wants in a teacher." *Scholastic*, Vol. xxviii, 1936, p. 9.
11. Reed, Anna Y.: *The Effective and Ineffective College Teacher.* New York: The American Book Company, 1935, 344 pp.
12. Reinhardt, Emma: "Personality traits in the teaching profession as analyzed by ninth-grade students." *School and Community*, Vol. xxiii, 1937, p. 43.
13. ———: "Do persons in later years rate their teachers as they did while in classes?" *Educational Administration and Supervision*, Vol. xxii, 1936, pp. 225–232.
14. Thorndike, E. L.: "The correlation between interests and abilities in college courses." *Psychological Review*, Vol. xxviii, 1921, pp. 374–376.

A CRITICAL COMPARISON BETWEEN THE SOLUTIONS OF THE FACTOR PROBLEM OFFERED BY SPEARMAN AND THURSTONE

HATTIE BAITY HEDMAN

University of Toronto

INTRODUCTORY STATEMENT

This paper* examines critically some of the major claims set forth in factorial analysis as applied to psychology. The discussion is based largely on the work of Spearman and of Thurstone who are identified with two apparently distinct solutions of the factor problem. It is believed that a comparison of the basic principles of these two authorities enables us, by exposing likenesses and differences, to clarify the "factor" concept and to evaluate the major claims currently set forth. The two systems are so deeply rooted in mathematical principles that neither can be fully appreciated without an understanding of the underlying mathematical logic. While from a systematic point of view statistical technique is secondary to method and, therefore, to psychological interpretation, there is, none the less, an interdependence between technique and psychological interpretation. It is believed, therefore, that an analysis and comparison of the techniques employed by these two authorities expose best their likenesses and differences in basic principles and enable us to reach a more definite, a more tangible notion of the psychological concepts involved. This paper accordingly deals largely with an analysis of techniques.

HISTORICAL REVIEW

Galton's Contribution.—Historically the factor approach is traced to the work of Galton, who may be said to have launched the movement by means of his pattern of thinking generally and by his correlational calculus in particular. Basic to factor analysis is the principle of bivariate correlation. And basic to a comparison of the techniques of Spearman and of Thurstone is the regression equation in a bivariate situation. For this equation we are indebted to Galton, who, through regression, reached the full significance of the "index of correlation" as a fraction for measuring the closeness of relationship. The equation

* An abstract of a thesis accepted by the University of Toronto for the degree of Doctor of Philosophy.

was originally developed by him for the purpose of measuring the resemblance of offspring to parent in respect to various characters. Its logic was then seen to apply as well in measuring the relationship between characters in the same individual. The ideas of the correlational calculus were opening a new world of thought to Galton. In 1883 (1, pp. 20–23) we find him definitely suggesting a relationship between marked ability and sensory discrimination. Thus, the first hint of factor analysis appears to have come from Galton.

Spearman's Contribution.—Spearman, inspired by Galton's *Inquiries into Human Faculty*, wished to find out whether, as Galton had indicated, the abilities commonly taken to be intellectual have any correlation either with each other or with sensory discrimination. The results from experiments which he was carrying on in a small village school in England indicated that the mental powers did correlate with each other in considerable degree (7, p. 287). Their magnitudes were also noticed to arrange themselves in an interesting systematic order. This at the time was said to be "hierarchical." The more precise term "equiproportionality" was introduced later.

Spearman was puzzled over the results obtained by Cattell, who, almost at the same time, was carrying on similar experiments in the Columbia University laboratory independently. The Columbia results were almost diametrically opposed to Spearman's. To test the accuracy of his own work Spearman repeated his experiments and found that the repetition always produced a value somewhat different from before. He came to realize that "the result actually obtained in any laboratory test must necessarily have been in every case perturbed by various contingencies which have nothing to do with the subject's real general capacity."[7] These errors of observation had always been ignored apparently on the assumption that half of them tend to enhance the correlation and half to reduce it. This is not the case. Errors of observation, while they tend to compensate each other in the matter of averages, tend to decrease, or attenuate, a correlation coefficient. The proof of this assertion has been given elsewhere in simple mathematical equations.[4] Examination of the Columbia research[17] showed that conditions conducive to error prevailed. The object, to be sure, is the accurate determination of the degree of correlation between abilities. But no technique for such determination was at hand. Further statistical analysis was not only justified, but was necessary. We are indebted to Spearman for the desired technique. By refining and extending the correlation techniques devised

by Galton he evolved a formula for eliminating the attenuating effect of observational errors in observed correlation. One form of correction presented by Spearman and widely used for two independent series of observations of each ability is the following:

$$r_{xy} = \frac{(r_{x_1y_1}r_{x_2y_2}r_{x_1y_2}r_{x_2y_1})^{1/4}}{(r_{x_1x_2}r_{y_1y_2})^{1/2}} \qquad \text{(10, Appendix } i\text{)}$$

What Spearman really wanted to know was the correlation between intelligence in general and sensory discrimination in general. Applying the corrective technique in suitable form to his data he found, through a period of research, that correlation to be practically perfect. To him this result indicated that the "common and essential element in the Intelligences wholly coincides with the common and essential element in the Sensory Functions." After further experimentation with his own and other data, Spearman concluded from the many observed facts that "all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (7, p. 284). He called the general factor "g" and the specific factor (respective error) "s." The true correlation is not a theoretical abstraction. It represents the limit to which the observed correlation will continually approach as improvement is made in the experimental procedure.

In 1912 Spearman and Hart[12] showed that when a general factor is present in a table of correlation coefficients, *i.e.*, when every variable can be divided into two factors g and s, the tetrad equation is necessarily satisfied. In other words, if the correlation obtained by the above corrective formula is unity we may write

$$r_{x_1y_1}r_{x_2y_2} = r_{x_1y_2}r_{x_2y_1} = r_{x_1x_2}r_{y_1y_2}.$$

Spearman's solution for the reverse problem; namely, that when the tetrad criterion is satisfied every variable would be divided necessarily into the two factors g and s, was given in 1922.[11] This precise and convenient criterion for a general factor, as soon as a formula for determining its reliability[6] could be developed, was used instead of the makeshift technique of intercolumnar correlation, or of the crude practice of judging hierarchical arrangement. The term "hierarchy" took the more precise meaning of "equiproportionality." Basic to

the solution of the inverse problem is the correlation between sums, demonstrated by Spearman in 1913,[9] the principle of which plays an important part in factor analysis.

Thurstone's Contribution.—Thurstone attempts to go beyond the practical limitations of Spearman's approach, or at least to supplement that approach. Spearman concentrated upon the notion of identifying one common factor at a time. The two-factor technique (perhaps better called the single factor technique) was the inevitable result. Thurstone's interest in manifold motives, needs, or important functional characteristics of the individual expressed itself in a desire to analyze human behavior generally in such a way as to show wherein the significant foci of individual differences lay. Intelligence, g, or any other single factor may be of great importance as an aspect of human nature, a dimension of personality. But what other analogous factors may we postulate as equally legitimate and convenient dimensions? Perhaps also what appears to be "g" in many of Spearman's investigations may need refinement by further factorial analysis.

Accordingly the movement known as "Multiple Factor Analysis" is launched.* Instead of selecting experimental situations giving data that reflect one common factor only, can we use data gathered from a wider, less carefully selected variety of situations, and determine the least number of common factors that need to be postulated, in order to account for the interrelationships demonstrated by the data?

It will be seen that the two approaches, Spearman's and Thurstone's, are not in any sense conflicting, although methodologically they are distinguishable. Many of the differences between them arise from essential differences in outlook. Their techniques are far more similar in statistical bases than appears at first sight. It is now proposed to examine critically those techniques. Since a good deal of critical evaluation of Spearman's work has already been published, and since Thurstone's formulation claims to be more comprehensive,

---

* In terms of temporal priority, the idea of a multiple factor technique was perhaps given statistical consideration first by Kelley.[3] Kelley seems, however, to have had a less clearly defined psychological outlook. Further, in the *Crossroads* (1929), while he argues for what later became known as multiple factor analysis, he uses a technique which really differs less from Spearman's than it does from Thurstone's. His subsequent work on multiple factors undoubtedly owes a great deal to Thurstone's formulation of the problem, and, statistically, to the solution proposed by Hotelling.[2] Because of the greater interest psychologically of the work of Spearman and of Thurstone, the present paper deals with those two authors exclusively.

we shall examine Thurstone's contribution, and relate it to the earlier single factor techniques in the process.

MATHEMATICAL COMPARISON—SPEARMAN AND THURSTONE

Thurstone's Basic Equation.—In his first contribution[13] to the factor problem, which was in 1931, Thurstone stated a basic equation, which he has used consistently to the present time. Expressed in standard scores, that equation is,
for test a

$$S_a = a_1x_1 + a_2x_2 + a_3x_3 \qquad \text{(i)}$$

for test b

$$S_b = b_1x_1 + b_2x_2 + b_3x_3 \qquad \text{(ii)}$$

where

S_a denotes the standard score of an individual in test a
x_1 denotes the standard score of the individual in ability No. 1
x_2 denotes the standard score of the individual in ability No. 2
x_3 denotes the standard score of the individual in ability No. 3
a_1 denotes the loading of ability No. 1 in test a
a_2 denotes the loading of ability No. 2 in test a
a_3 denotes the loading of ability No. 3 in test a
S_b denotes the standard score of the individual in test b
b_1 denotes the loading of ability No. 1 in test b
b_2 denotes the loading of ability No. 2 in test b
b_3 denotes the loading of ability No. 3 in test b

From the two applications of the postulated basic equation above, equations (i) and (ii), and on the assumption that the three abilities are uncorrelated, the correlation between tests a and b is expressed as follows:

$$r_{ab} = \frac{\Sigma S_a S_b}{N} = a_1b_1 + a_2b_2 + a_3b_3 \qquad \text{(iii)}$$

And the standard deviation of a series of standard scores, say, S_a, is

$$\sigma^2_{(S_a)} = \frac{\Sigma(S_a)^2}{N} = a_1^2 + a_2^2 + a_3^2 = 1 \qquad \text{(iv)}$$

What Thurstone desired was a formula, in terms of obtainable correlation, for estimating the first factor loadings, based on the

assumption that the (psychological) "abilities" may be identified with (statistical) "factors." He appears, however, not to have offered a direct algebraic solution for such a formula. What he did was to adopt a geometrical interpretation of correlation by considering equation (iv) as the equation of a sphere with a_1, a_2, a_3 as the coördinates of test a considered as a point on the surface of the sphere. The coördinates of the centre of gravity of a set of tests or points, say, s points, are obviously

$$\frac{\Sigma s_1}{N_s}, \quad \frac{\Sigma s_2}{N_s}, \quad \frac{\Sigma s_3}{N_s}$$

By allowing the x-axis to pass through the centre of gravity of the set of s points the two other coördinates, say, y and z, must be zero. Therefore

$$\Sigma s_2 = \Sigma s_3 = 0$$

Analogous to equation (iii)

$$\begin{aligned} r_{as} &= a_1 s_1 + a_2 s_2 + a_3 s_3 \\ \Sigma r_{as} &= a_1 \Sigma s_1 + a_2 \Sigma s_2 + a_3 \Sigma s_3 \\ &= a_1 \Sigma s_1 \\ \Sigma r_{ss} &= (\Sigma s_1)^2 \\ a_1 &= \frac{\Sigma r_{as}}{\Sigma s_1} = \frac{\Sigma r_{as}}{\sqrt{\Sigma r_{ss}}} \end{aligned} \qquad \text{(v)}$$

Let it be noted that when Thurstone fixes the x-axis through the centre of gravity of the s tests and thereby reduces the y and z coördinates to zero, he automatically eliminates the second and third factor loadings. Is not this equivalent to asking simply for a first factor loading, *i.e.*, of writing $S_a = a_1 x_1$? It is interesting to note that by the nature of the centre of gravity method the factor loadings other than the first have been eliminated by "pooling" the series of s tests, taking the mean, and then measuring a particular test by this mean value to get the first factor loading of that test. This suggestion leads directly to a discussion of equation (v) and a strict examination of it algebraically in order to clarify it.

ANALYSIS OF THE FORMULA FOR FIRST FACTOR LOADING

The equation for the loading of the first factor, as stated above (equation v), is

$$a_1 = \frac{\Sigma r_{as}}{\sqrt{\Sigma r_{ss}}}$$

where a_1 denotes the loading of factor one in test a, Σr_{as} is the sum of the correlation between a and every test in group s, including a, and s represents that group of tests in the whole table of correlation coefficients under consideration which correlate positively with test a, plus test a itself. Test a is that test which has to its credit the highest arithmetical sum of correlation coefficients in the whole table. The Σr_{ss} of the denominator is twice the algebraic sum of the correlation between every s test and every other s test, plus the self correlations.* It is, therefore, the algebraic sum of the whole table of correlation coefficients for the s group. This practice of using subgroups was abandoned by Thurstone in later work. The whole table of correlation coefficients came to be used instead of the selected s group. Otherwise the procedure remained the same.

If, at this point, we compare Thurstone's analysis for multiple factors with Spearman's analysis for a general factor we will find that Spearman has derived and applied, as related to his general factor theory, an equation identified with Thurstone's expression for first factor loadings when all correlations are positive. This equation was given by Spearman (9, p. 419; 10, app. iv; 5, p. 216) in the following form:

$$r_{ag} = \frac{\sqrt{M}\bar{r}_{ax'}}{\sqrt{1 + (M-1)r_{x'x''}}}$$

where each standard deviation has been reduced to unity, as a result of which the regression coefficients have become equal to each other and to the coefficient of correlation, and deviations have become standard scores; where r_{ag} expresses the correlation between a and the sum of the x tests which include a, or the mean of the sum; where M denotes the number of variables in the x table, and $\bar{r}_{ax'}$ is the mean correlation between a and every test in the x group which includes a. In the denominator $r_{x'x''}$ is the mean of all the correlation coefficients in the x table not including self correlations. Each coefficient is used twice. The self correlation is accounted for by the numeral one.

* Originally, Thurstone assumed all self correlations to be unity. He later, in his centre of gravity method, accepted the "highest correlation coefficient in the column" as a more legitimate value for reasons which are obviously related to his change in method.

Let us reduce this expression thus:

$$r_{ag} = \frac{\sqrt{M}\bar{r}_{ax'}}{\sqrt{1 + (M-1)\bar{r}_{x'x''}}}$$
$$= \frac{\Sigma r_{ax'}}{\sqrt{M + \Sigma r_{x'x''}}}$$

The expression for r_{ag} is now seen to be directly comparable to the Thurstone expression

$$a_1 = \frac{\Sigma r_{as}}{\sqrt{\Sigma r_{ss}}}$$

for both r_{ag} and a_1 are equal to $r_{a:\Sigma s}$ including a. In brief, the loading of test a with factor one in a group of s tests is equivalent to the correlation between a and the sum of the group of tests including a; or, otherwise expressed, is equal to $r_{(a):(s_1+s_2+\ldots+s_n)}$, all tests having been expressed in standard scores. Hence Thurstone's first factor loading of a test or Spearman's r_{ag} expresses the correlation between that test and the composite score of a chosen battery,* both variables expressed in standard scores.

Let it be noted that the correlation between a variable and the sum of, *i.e.*, the composite variable derived from a series of variables is the same thing as the correlation between a variable and the mean variable derived from the series of variables (5, p. 222). Therefore, we may write

$$a_1 = r_{(a):(s_1+s_2+\ldots+s_n)} = r_{(a):(\text{mean } s)}$$
$$r_{ag} = r_{(a):(x_1+x_2+\ldots+x_n)} = r_{(a):(\text{mean } x)}$$

MEANING OF "LOADING"

(*a*) *In Terms of Regression.*—Spearman's equation

$$r_{ag} = \frac{\Sigma r_{ax'}}{\sqrt{M + \Sigma r_{x'x''}}} = r_{a:\Sigma x}$$

or Thurstone's equivalent

$$a_1 = \frac{\Sigma r_{as}}{\sqrt{\Sigma r_{ss}}} = r_{a:\Sigma s}$$

* It should be noted, however, that the calculation of r_{ag} by the formula $(r_{ag} = r_{ab}r_{ac}/r_{bc})^{1/2}$, given by Spearman in the *Abilities of Man*, is independent of the self correlations, which, when used, are given assumed values.

indicates the loading of test a with the communality of the battery of tests under analysis. If, as in the case of the Spearman procedure, the battery is indicative of intellectual ability, then our equation would indicate the extent to which a is loaded with g. If, for example, we let

$$r_{ag} = .8$$

then we may write the following regression equation:

$$\bar{a} = r_{ag} \cdot g = .8g \text{ standard scores}$$

where r_{ag} has the meaning given immediately above, where g is the standard score of an individual in the mean series derived from the group of x tests, and where $\bar{a}$ is the most probable standard score of the individual in test a predicted from r_{ag} and g, *i.e.*, $\bar{a}$ is the mean standard score in test a of all individuals of the given g value and is, therefore, located geometrically on the line expressing the regression of a on g. Accordingly, in the case of our example, the most probable standard score of an individual in test a is .8 times his standard score in the mean series, *i.e.*, the g series, derived from the x measures. That is the precise use of "loading." That notion applies as well to Thurstone's concept with his wider range of batteries, since his analysis comes to be based on two variables and, therefore, on the laws governing the relationship between two variables. A certain deviation in an individual's measure in factor one indicates some deviation in his measure in test a, depending upon the loading of test a with factor one. These considerations indicate that "loading" may be rather misleading, and needs to be interpreted with care. Misrepresentation may be avoided by keeping in mind the fact that the term "loading" is, as already indicated, simply a coefficient of correlation* expressing the relationship between two variables which have been reduced to standard scores. The one variable, in the case under discussion—that of the first factor loading of a test—is that test; while the other variable is the composite series, or the mean series, derived from a battery of tests.

(*b*) *As Distinguished from a General Factor.*—In any table of intercorrelations loadings of factor one may be determined for each variable by Thurstone's technique. But according to the meaning logically

* Or a regression coefficient. Since the variables are all expressed in standard scores the regression coefficients are equal to each other and to the correlation coefficient.

assigned to "loading" this does not demand that the loadings (a, b, . . . , z) refer to a general factor, *i.e.*, refer to a factor common to every variable of the battery under analysis. Logically, a general factor can only be implied when a one to one relationship appears after the intercorrelations have been corrected for the attenuating effect of errors and specific factors. But, when a one to one relationship prevails, the formula for correction for attenuation demands that the tetrad difference be zero.[12] Thus functional identity demands that the tetrad equation be satisfied. Conversely, when the tetrad equation is satisfied every variable is divided necessarily into the two factors g and s.[11] It is *imperative*, therefore, that in a table of intercorrelations Thurstone's factor one will be general and common to all the variables *only* when the tetrad criterion is satisfied. Factor loadings may be very misleading unless the tetrad criterion is satisfied, *i.e.*, unless the influence of the factors other than the common one is examined.

APPLICATION OF THE LAW OF REGRESSION

(*a*) *To the Basic Equation.*—At this point it is well to subject to precise analysis the initial equation from which, through a geometrical interpretation, the formula for first factor loading was derived. That equation, with symbols as previously defined, is

$$S_a = a_1x_1 + a_2x_2 + a_3x_3 \qquad \text{(i)}$$

Here, it is assumed that the factors combine in a linear manner in order to produce the total performance of a test for an individual. Let us compare the above equations with the usual regression equation. We may approach the comparison by setting up the following equation:

$$A = r_{ax} \cdot X + D_a' \qquad \text{(ii)}$$

where A and X are expressed in units which render their standard deviations equal, where A indicates the gross score of an individual in test a, where X indicates the gross score of the individual in the mean series derived by combining a group of x tests which are characteristic of the first factor. It is obvious that r_{ax} corresponds to a_1. We let D_a' denote that part of measurement A which is not accounted for by $r_{ax} \cdot X$. It is specific in the sense of being uncorrelated with X. Then

$$\frac{\Sigma A}{N} = \frac{r_{ax}\Sigma X}{N} + \frac{\Sigma D_a'}{N}$$

$$A_m = r_{ax} \cdot X_m + D'_{a_m} \qquad \text{(iii)}$$

From equations (ii) and (iii)

$$A - A_m = r_{ax}(X - X_m) + D_a' - D'_{a_m}$$

$$a = r_{ax} \cdot x + d_a' \text{ (deviations in points)} \qquad \text{(iv)}$$

or

$$a = a_1 x + d_a' \qquad \text{(v)}$$

where d_a' indicates the addition to $r_{ax} \cdot x$, or to $a_1 x$, that is necessary to account for a. It is obvious that the standard deviations of a and x are equal to those of A and X, respectively, and, therefore, to each other. The standard deviation of d_a' is equal to that of D_a'. Using equation (iv) let us reduce d_a' to a value the standard deviation of which is equal to that of the variables a and x. We will then have the standard deviations of the variables A, X, a, x, converted D_a' and converted d_a' all equal. Thus

$$d_a' = a - r_{ax} \cdot x$$

$$\Sigma(d_a')^2 = \Sigma a^2 - 2r_{ax}\Sigma ax + r_{ax}^2 \Sigma x^2$$

$$\sigma(d_a') = \sigma_a(1 - r_{ax}^2)^{\frac{1}{2}} \qquad \text{(vi)}$$

which is the well known error of estimate, and indicates here that d_a', as well as D_a', ranges practically between $\pm 3\sigma_a(1 - r_{ax}^2)^{\frac{1}{2}}$ But

$$\frac{'d_a'}{\sigma(d_a')} \cdot \sigma_a = (say)da$$

which is a deviation in the variable D_a corresponding to d_a' in the variable D_a'. Since the standard deviation of D_a is equal to that of A

$$\frac{d_a'}{\sigma_a(1 - r^2{}_{ax})^{\frac{1}{2}}} \cdot \sigma_a = d_a$$

$$d_a' = d_a(1 - r_{ax}^2)^{\frac{1}{2}} \qquad \text{(vii)}$$

We may now write for equations (iv) and (v)

$$a = r_{ax} \cdot x + d_a(1 - r_{ax}^2)^{\frac{1}{2}} \qquad \text{(viii)}$$

$$a = a_1 x + d_a(1 - r_{ax}^2)^{\frac{1}{2}} \qquad \text{(ix)}$$

But by partial correlation the above equations may be expressed in terms of $r_a d_a$ as follows:

$$r_{ad_a} \cdot x = \frac{r_{ad_a} - r_{ax} \cdot r_{d_a x}}{(1 - r_{ax}^2)^{1/2}(1 - r^2_{d_a x})^{1/2}} = 1 \qquad \text{(x)}$$

Clearly

$$r_{ad_a} \cdot x = 1 \qquad \text{and} \qquad r_{d_a x} = 0$$

Therefore from equation (x)

$$r_{ad_a} = (1 - r^2_{ax})^{1/2} \qquad \text{and} \qquad r_{ax} = (1 - r^2_{ad_a})^{1/2}$$

And from equation (vii)

$$d_a' = d_a r_{ad_a}$$

This enables us to write for equations (iv), (v), and (ii)

$$a = r_{ax} \cdot x + r_{ad_a} \cdot d_a$$

$$a = a_1 x + r_{ad_a} \cdot d_a \qquad \text{(xi)}$$

$$A = r_{ax} \cdot X + r_{ad_a} \cdot D_a \qquad \text{(xii)}$$

Or expressed in the usual regression style

$$\bar{a} = r_{ax} \cdot x$$

$$\bar{a} = a_1 x$$

It is to be noted that $r_{ad_a} \cdot d_a$ with its probable range indicates the distance a is from the regression line of average values designated $\bar{a}$. It is equivalent to $a - \bar{a}$. This operation is now legitimate since a, x, and d_a have equal standard deviations and are, therefore, additive statistically. It is also to be noted that in equation (xi) what we add to the regression of A on X is the regression of A on D_a (errors plus specific factors). Let us now compare equations (i) and (xii) after having expressed the latter in standard scores.

$$\frac{a}{\sigma_a} = \frac{a_1 x}{\sigma_a} + \frac{r_{ad_a} \cdot d_a}{\sigma_a} \qquad \text{(xii)}$$

$$S_a = a_1 x_1 + a_2 x_2 + a_3 x_3 \qquad \text{(i)}$$

The left members are identical. The first terms in the right members are identical.

Then $a_2 x_2 + a_3 x_3$ should equal $r_{ad_a} \cdot d_a / \sigma_a$ which, when expressed in the original units, is equal to d_a' with a standard deviation of

$\sigma_a(1 - r^2_{ax})^{\frac{1}{2}}$; and, therefore, falls practically within the range of $\pm 3\sigma_a(1 - r^2_{ax})^{\frac{1}{2}}$, *i.e.*, within the range of the error of estimate.

It is difficult, therefore, to accept Thurstone's basic equation; namely, that $S_a = a_1x_1 + a_2x_2 + a_3x_3$, unless it can be shown that the addition of $a_2x_2 + a_3x_3$ comes within the bivariate regression law; since, as shown above, a_1x_1 accounts for the whole of *Sa* apart from the error of estimate.

Thurstone's contribution of 1931 was followed by several studies in which he produced more elaborate techniques in the solution of the multiple factor problem. The method of principal axes in a simple form is given in the *Theory of Multiple Factors* published in 1933. The centroid or centre of gravity method is Thurstone's *Simplified Multiple Factor Method* published a little later in 1933. *The Vectors of Mind*, which appeared in 1935, summarizes his multiple factor methods to date. An examination of the techniques of these more recent contributions shows that Thurstone has consistently used his original basic equation, which we have shown to be of questionable value. Our claim; namely, that in a bivariate situation one common factor plus the error of estimate is sufficient to explain an individual standard score holds throughout our critical estimate of Thurstone's entire work. He has never altered the situation.* He has persisted in using a technique which was designed for taking only one factor at a time.

(*b*) *The Question of Residuals.*—Thurstone attempted to meet the difficulty in obtaining more than one factor at a time by introducing the method of computing residuals. A residual represents the discrepancy between the obtained coefficient of correlation and the coefficient computed from $r'_{ab} = a_1b_1$. Thus $\rho_{ab} = r_{ab} - r'_{ab}$. The residual ρ_{ab} is the difference between two bivariate correlation coefficients expressing the relationship between variables *a* and *b*. This difference is not itself a coefficient which we can define. We would question Thurstone's use of it as a correlation coefficient.

Residuals, we claim, are due to the assumptions made in deriving equation $r'_{ab} = a_1b_1$ from two applications of the basic equation, such as

$$S_a = a_1x_1 + e_a + s_a' \text{ standard scores}$$
$$S_b = b_1x_1 + e_b + s_b' \text{ standard scores}$$

---

* If the basic equation is extended to include errors of measurement (e_a) and specific factors (s_a') the above criticism still holds.

or, expressed in the language of this study and according to the logic of the regression equation, such as

$$a = r_{ax} \cdot x + r_{ad_a} \cdot d_a$$
$$b = r_{bx} \cdot x + r_{bd_b} \cdot d_b$$

From either set of equations we may derive

$$r'_{ab} = a_1 b_1$$

based on the assumption of uncorrelated factors. As already indicated, the addition of $r_{ad_a} \cdot d_a$ with its probable error is a means of making up the difference between the actual measure a and the mean measure $\bar{a}$ predicted in the regression equation $\bar{a} = r_{ax} \cdot x$. We might view this departure from the regression line of typical values as due to errors or to special ability or to both.*

SUMMARY AND CONCLUSIONS

Historically the factor approach is traced to the work of Galton, who may be said to have launched the movement by means of his pattern of thinking generally and by his correlational calculus in particular. Thence the subsequent contributions of Spearman are examined in some detail.

The techniques of Spearman and Thurstone are scrutinized, and a somewhat detailed comparison is undertaken. These authors appear at first sight to be identified with two distinct statistical solutions to the factor problem. The basic principles employed, however, are shown to be identical, in that Spearman's formula for determining the "correlation of sums" and Thurstone's method of extracting "first factor loadings" are shown to be identical.

Further analysis of the basic equation used by Thurstone reveals the questionable nature, from a statistical point of view, of his method of extracting more than one general factor at a time. In this analysis

---

* It has been suggested by some critics that the above argument against the treatment of residuals would invalidate all use of partial correlation. This is far from true. The method of partial correlation is not the same as that of taking residuals as practiced by Thurstone. This is equivalent to saying that partial correlation is not based on the subtraction of bivariate correlation coefficients. In the case of partial correlation three or more variables are used, and regression equations are set up accordingly. The partial regression coefficients are reduced to bivariate or obtainable values from which the formula for estimating partial correlation is readily derived. Yule's formula has been analyzed elsewhere[4] in a manner that can be appreciated by those unacquainted with the higher mathematics.

the application of the regression equation shows that one general factor is sufficient to explain an individual standard score in a bivariate situation. The Thurstone technique is inadequate for the method proposed by him; namely, that of extracting multiple factors. The technique belongs in the method where one general factor at a time is extracted such as the method of Spearman from which it was derived, and is less adequate for that purpose than is Spearman's.

Further, the practice of taking residuals—the difference between two correlation coefficients—and then treating them themselves as correlation coefficients is questionable.

The force of this criticism of Thurstone's multiple factor techniques applies equally to all the methods so far proposed by him, such as the subgroups method, the method of principal axes, and the centroid method.

BIBLIOGRAPHY

1. Galton, Francis: *Inquiries into Human Faculty.* J. M. Dent and Son, London.
2. Hotelling, Harold: "Analysis of a complex of statistical variables into principal components." *J. of Educ. Psych.*, Vol. XXIV, Nos. 6 and 7, Sept. and Oct., 1933, pp. 417–441, 498–520.
3. Kelley, Truman L.: *Crossroads in the Mind of Man.* Stanford University Press, 1928.
4. Line, W. and Hedman, H. B.: "A simplified statement of the two factor theory." *J. of Educ. Psych.*, Vol. XXIV, No. 3, March, 1933, pp. 195–219.
5. ———: "The reversibility of proof." *J. Exper. Educ.*, Vol. III, 1935, pp. 218–223.
6. Pearson and Filon: "On the probable error of frequency curves." *Philos. Trans.*, Vol. CXCI, A, 1898, p. 229.
7. Spearman, C.: "General intelligence objectively determined and measured." *Am. J. Psych.*, Vol. XV, 1904, p. 201.
8. ———: "The measurement of association between two things." *Am. J. Psych.*, Vol. XV, 1904, p. 72.
9. ———: "Correlation of sums and differences." *Brit. J. Psych.*, Vol. V, 1913, pp. 417–426.
10. ———: *The Abilities of Man.* Macmillan and Co., Ltd., 1927.
11. ———: "Correlation between arrays in a table of correlations." *Proc. Roy. Soc.*, A, 1922, p. 101.
12. Spearman and Hart: "General ability: its existence and nature," *Brit. J. Psych.*, Vol. V, 1922, p. 51.
13. Thurstone, L. L.: "Multiple factor analysis." *Psych. Rev.*, Vol. XXXVIII, 1931, pp. 406–427.
14. ———: *Theory of Multiple Factors.* Edwards Bros., 1933.
15. ———: *The Vectors of Mind.* University of Chicago Press, 1935.
16. ———: *A Simplified Multiple Factor Method.* The University of Chicago Bookstore.
17. Wissler, Clark: *Psych. Mon.*, Vol. III, No. 6, 1901.

AN EVALUATION OF A SCOUT LEADERS TRAINING PROGRAM

HARRY K. EBY

Boy Scouts of America

AND

FORREST A. KINGSBURY

University of Chicago

The formal training program for volunteer scoutmasters carried on by the Boy Scouts of America includes the following courses: First, an elementary course of twelve sessions, each two to five hours in length, dealing with the essentials of scouting objectives, organization, and methods, divided into two parts, of six lessons each, leading respectively to the "Minimum" and the "Approved" Certificates; second, an advanced course of the same length, the "Principles of Scoutmastership"; and third, a number of specialization courses, such as First Aid, Patrol Leadership Training, Troop Camping, Indian Lore, etc. Parallel courses are provided for skippers and cubmasters, in the Sea Scout and Cubbing departments. Instruction is predominantly practical and active, rather than verbal, although it includes reading assignments, talks, and discussions.

During 1937 an effort was made to evaluate this training as carried on in the Chicago Council, with particular reference to the Minimum course, which is prerequisite to the Approved and Principles courses, and is taken by more men than any other one course. The rated efficiency of the troops, ships, and packs was studied in relation to their leadership by men who had completed and men who had not completed this course. Under the direction of Harry K. Eby, then Assistant Executive of Chicago Council in charge of education and organization, now Assistant Director of Education and Relationships in the national organization, training data for sixteen hundred forty-four volunteer commissioned leaders and assistants—eight hundred twenty-five scoutmasters, skippers, and cubmasters, and eight hundred nineteen assistant scoutmasters, mates, and assistant cubmasters—were compared with the achievement index ratings of the units with which they were identified as of January 1, 1937.

The Chicago Council's achievement index is a method of assigning quarterly and annual point-ratings to each unit, whether troop,

ship, or pack, on the basis of five general lines of achievement; namely, Individual scout advancement, Outdoor program activities, Leader participation in leaders' meetings and training courses, Troop participation in community activities, and Membership growth and tenure. Units are rated *A*, *B*, *C*, which are regarded as satisfactory; *D*, or *E*, which are unsatisfactory. Frequency distribution of these five grades does not conform to the normal curve because of inequalities in the spread of achievement indices for the several grades; intermediate grades *B*, *C* and *D* are narrow in spread and show lower frequency, while *A* and *E* ranges are much wider and more frequent. Several types of comparison between the two variables—training and achievement index—were made.

(1) The first comparison (Fig. 1) shows the percentages of the eight hundred twenty-five leaders who had completed the Minimum course to be as follows: In grade *A* units, seventy-five per cent; *B*, sixty-nine per cent; *C*, seventy-one per cent; *D*, fifty-five per cent; *E*, thirty-seven per cent. The differences between these percentages and one hundred per cent gives the corresponding proportions of untrained leaders in the several grades. This amount of training, it will be observed, serves to differentiate the least efficient from the more efficient units better than it does between the three satisfactory grades. For assistant leaders (Fig. 1*b*) the relationship was considerably lower than among leaders, only the lowest—the *E* grade—showing a reliably lower percentage of training than the grades above it. This is hardly surprising considering that it is the scoutmaster and not his assistant who is normally responsible for the tone and quality of the troop. Consequently, most of the further comparisons were confined to the eight hundred twenty-five scoutmasters, skippers, and cubmasters. A second percentage comparison (Fig. 2) with units grouped into two classes, satisfactory and unsatisfactory, showed that of four hundred three leaders (omitting assistants) in satisfactory units, seventy-three per cent were trained and twenty-seven per cent untrained; while of four hundred twenty-two in unsatisfactory units, forty-four per cent were trained and fifty-six per cent untrained.

(2) The same general relationship appears in the correlation coefficients (tetrachoric r's) which were computed, between the satisfactory-unsatisfactory unit rating and the trained-untrained leader status, which proved to be +.47 for the eight hundred twenty-five leaders and only +.12 for the eight hundred nineteen assistant leaders.

(3) If we limit our comparisons to the two extreme groups of units (Fig. 3), *A* and *E*, the contrast is made somewhat sharper. Thus, in the *A*-grade units, seventy-five per cent of the two hundred forty-seven leaders were trained, and in the *E*-grade units, thirty-seven per

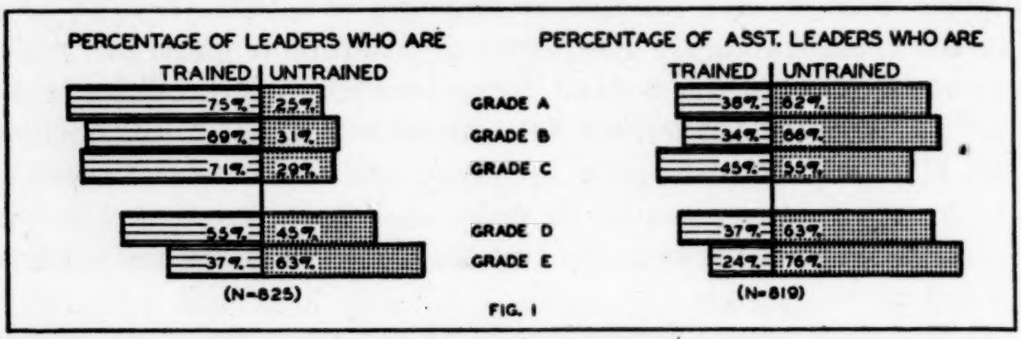


FIG. 1

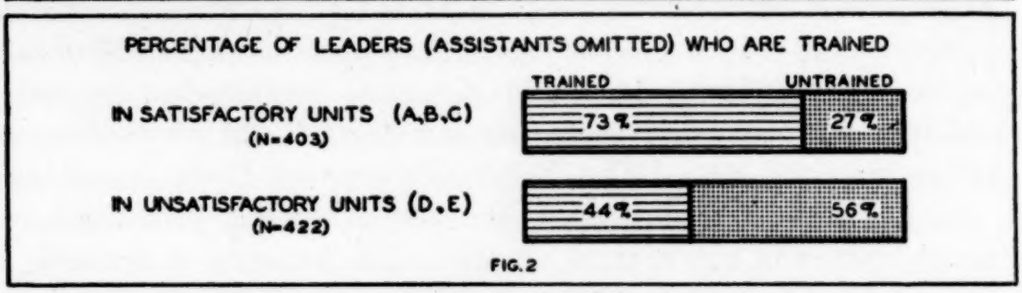


FIG. 2

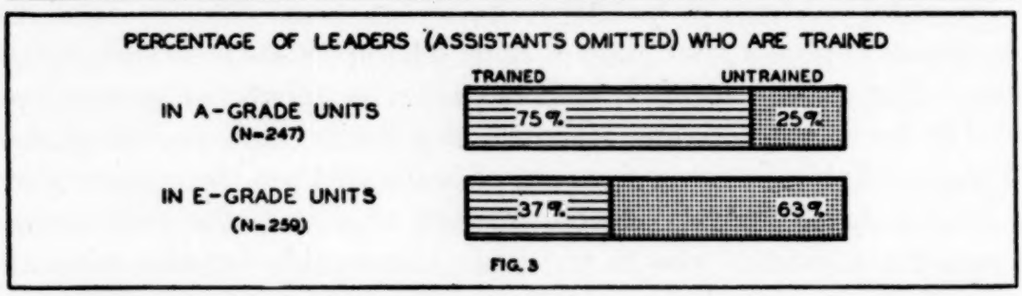


FIG. 3

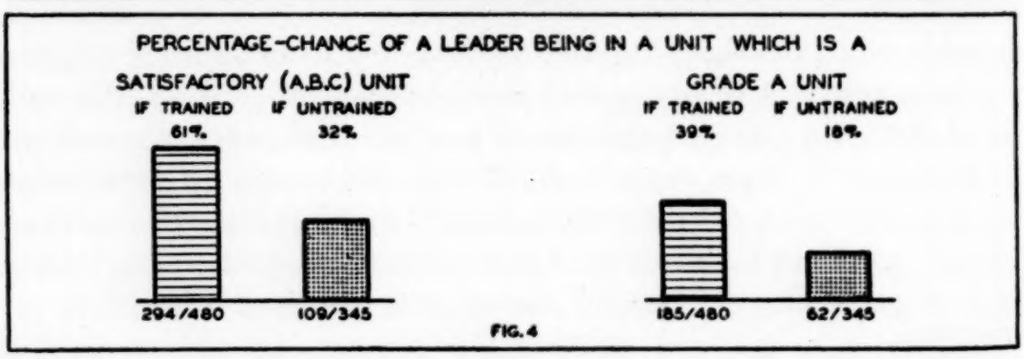


FIG. 4

cent of the two hundred fifty-nine leaders; while the corresponding percentages of untrained leaders were, *A*, twenty-five per cent; *E*, sixty-three per cent.

(4) Percentages can also be stated in terms of a leader's chances of being found either in a satisfactory unit or in an *A*-grade unit

(Fig. 4). An untrained leader was found to stand one hundred nine chances in three hundred forty-five of being in a satisfactory (*A*, *B* or *C*) unit, *i.e.*, thirty-two per cent; whereas a trained leader's chances

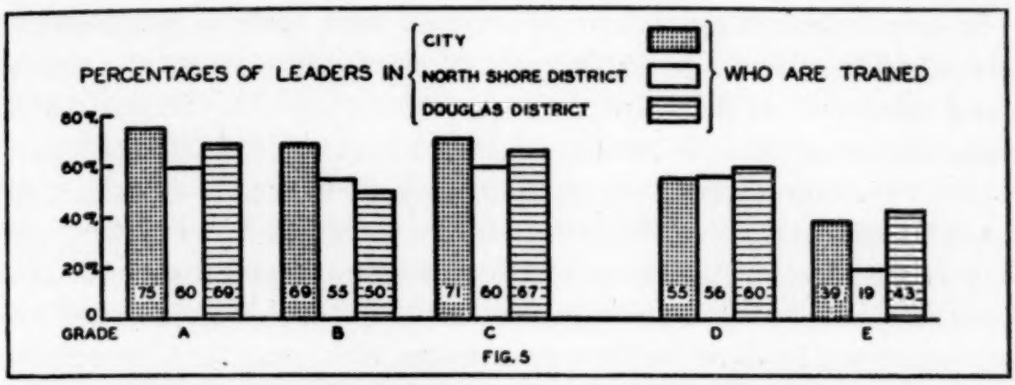


FIG. 5

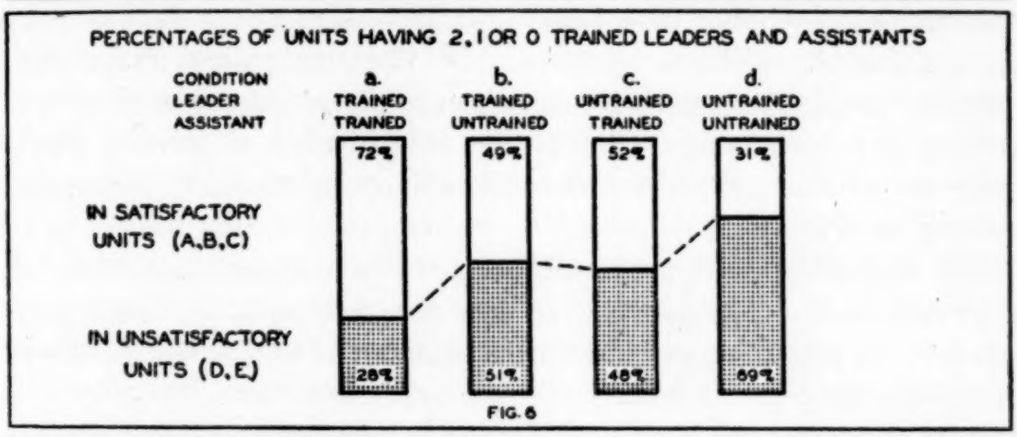


FIG. 6

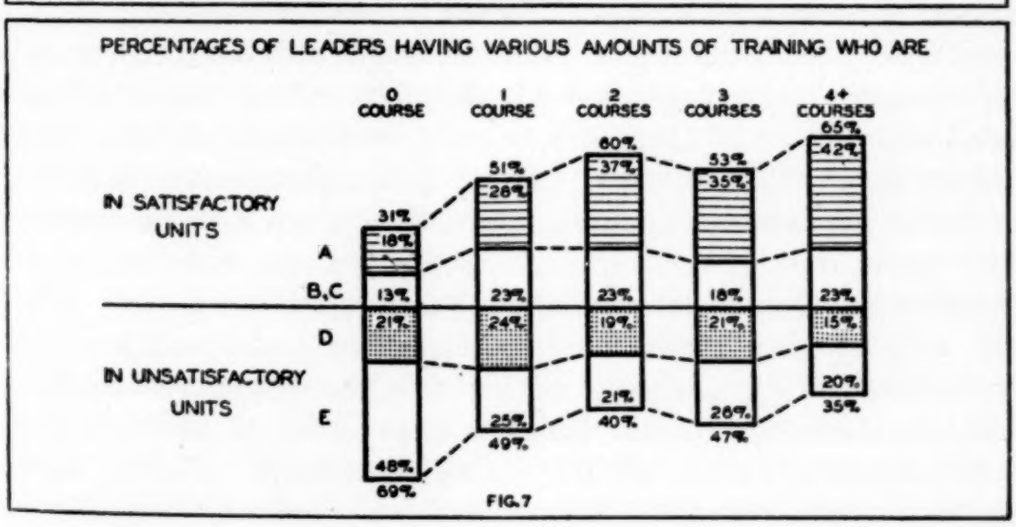


FIG. 7

were sixty-one per cent (two hundred ninety-four in four hundred eighty). In other words, the trained leader, even though training be taken to mean nothing more than the elementary course, has nearly

double the chance the untrained has of being in a satisfactory unit. If we limit our comparison to the highest, *A*-grade units, we find the corresponding probabilities are eighteen per cent for the untrained leader, and thirty-nine per cent for the trained, or more than double.

It seems clear from these comparisons that there is a definite and demonstrable advantage on the part of the leaders who have received even a minimum of formal training, so far as the achievement of their troops, ships, or packs is concerned. But that the elementary training course is the sole cause of this superiority does not, of course, follow. At least three other possibilities must be considered. The first possibility is that some units rate higher for other reasons such as better community advantages and support, more favorable human material, more educated leaders, better supervision, etc., and that leaders from these units may thus be stimulated to realize the value of formal training for themselves. Second, the training course may have a selective function, attracting men who are above the average in native capacity for leadership. Third, the relationship described may be partly an artifact, due to the achievement-index itself containing a training credit.

The first of these three possibilities receives some support so far as the presence of supplementary factors is concerned, although contradiction in its other aspect, from a comparison of two of the six districts into which the city is divided. The North Shore district undoubtedly ranks at least as high as any of the six in economic and social advantages, while the Douglas district, in the negro neighborhood, is probably the most underprivileged of all. The other four districts all contain a mixture of favored and unfavored neighborhoods, making analysis on this basis difficult. The two districts mentioned, however, are fairly homogeneous, and are practically equal with respect to achievement rating, ranking best among the six districts, whether measured in terms of frequency of *A* units (thirty-nine per cent for both, as compared with the city's thirty per cent), or fewness of *E* units (Douglas twenty-one per cent, North Shore twenty-two per cent, City thirty-one per cent). But this superiority of the two so dissimilar districts is achieved in quite different ways. North Shore, in spite of its high rating, has the smallest proportion of trained leaders of any district in the city (fifty per cent, as compared with the city-wide fifty-nine per cent); while Douglas, with sixty-one per cent trained has slightly more than the city average. It is apparent that North Shore's superiority must be due in more than ordinary measure

to some of the community factors mentioned, which compensate for the lower frequency of trained leaders. Douglas's high standing, on the other hand, is probably due in large measure to its leader training, since the other conditions mentioned seem wanting. This conclusion receives support from Fig. 5, in which a comparison is shown of the training frequencies in these two districts and in the city at large, grade by grade. The satisfactory grades, *A*, *B*, *C*, are attained in North Shore with fewer trained leaders than elsewhere, while Douglas requires more than North Shore. (The Douglas percentages are not so reliable as in other districts, because of small numbers involved, especially in the intermediate grades, the entire district totaling only forty-three units.) But that formal training (or lack of it) is, nevertheless, important in North Shore as well as elsewhere, is strikingly evident when we look at the poorest grade, the *E* units, which in North Shore have only half the proportion of trained leaders as the rest of the city; that is, twice as large a proportion of untrained leaders.

The North Shore district's condition, moreover, fails to support the suggestion that a social environment which makes for better-grade units will stimulate leaders to seek training. Our belief is thereby strengthened that high frequency of training of leaders is a cause, rather than an effect, of increased unit efficiency.

The second possibility, that the leaders who seek training average higher in native aptitude for leadership than those who do not seek training, can neither be conclusively refuted nor proved by the data at hand. A definite answer would require a survey of leaders, both trained and untrained, using various aptitude tests. But the district differences cited certainly point away from more than towards this possibility as a real factor; and, in the absence of positive evidence, it seems safe to disregard it.

The third possibility mentioned, that the relationship may be loaded with credit for training, is one which needs further study. This would involve recomputing the achievement-index of all units omitting that part of the training credit which overlaps the training variable; and this task has thus far not proved feasible. It has, however, been carefully considered by those familiar with the ratings, and on analysis does not appear to be a very important factor, for the following reasons:

(1) Elementary training courses completed in previous years will not appear in the calculation of the achievement index, which gives

credit for training courses only if completed during the period of the rating; (2) training courses in progress but not yet completed will affect neither variable; (3) an elementary course completed during the year can affect the index for only one quarter at most, and not the other three quarters included in the annual rating; (4) courses credited to troop committeemen, or to junior or assistant leaders, do not affect a leader's training status; (5) formal and informal training credit in the achievement index, except for the leader's one elementary course, does not affect his status as a trained leader in the comparisons previously cited; (6) the total amount by which the training credit can affect unit rating is at most only a small fraction of the total rating. This factor, then, can at best affect only a small proportion of the leaders and in only slight degree; so that its effect, in the few cases where it has any, would probably not materially alter the general picture.

Three other incidental questions were investigated:

(1) The absence of definite relationship between unit efficiency and training of assistant leaders seems to mean that the leader's training influences the unit's rating more than does the assistant's training. But it might possibly mean also that the trained assistants usually work under trained leaders, so that if a unit lacks a trained scoutmaster, skipper, or cubmaster, it is likely to have no trained leadership at all. To test the truth of this hypothesis, a comparison was made of the frequency and relation of the four possible combinations of trained and untrained leader, trained and untrained assistant. Of the eight hundred nineteen assistants, two hundred sixty-nine had training, and of these two hundred sixty-nine, seventy-one per cent were found to be associated with trained leaders, thus corroborating in some degree the hypothesis mentioned. The four possible combinations referred to revealed the following percentages of satisfactory units (Fig. 6):

	Per Cent
a. Units having both leader and assistant trained	72
b. Units having trained leader, untrained assistant	49
c. Units having untrained leader, trained assistant	52
d. Units having no trained leadership	31

The small difference, in reverse of expectation, between conditions *b* and *c* is doubtless statistically insignificant. In general, these results indicate a positive relationship between amount of trained leadership (two, one or no men) in a unit, and its index rating; but the

similarity between conditions *b* and *c* suggests that it is less important which one is trained, leader or assistant, than that one of them shall be trained.

One fact having a possible bearing on this low relationship is that most Assistant Scoutmasters have come up through scouting ranks, often serving as patrol leaders, and thus have had fairly recent experience with current scouting practices, methods, and responsibilities, whether formally trained as leaders or not. On the other hand, a larger proportion, of Scoutmasters have had no recent experience, or no experience at all, as scouts, so their scouting knowledge depends more on their formal training; although in the Chicago area over fifty per cent of the Scoutmasters have been scouts or Assistant Scoutmasters.

(2) A second question relates to the differences between the three departments of scouting. Scout troops greatly outnumber sea-scout ships and cub packs; they have been longer established, and their training programs more thoroughly worked out and tested; so that it is not surprising to find that the tetrachoric *r*'s between training and satisfactory rating run as follows:

629 Scoutmasters	$r_t = +.48$
60 Skippers	$r_t = +.47$
135 Cubmasters	$r_t = +.38$

The same general effect of the differences mentioned appears in comparisons of the percentage frequency of satisfactory units in the three departments:

50 per cent of the Scoutmasters are in satisfactory units;
48 per cent of the Skippers are in satisfactory units;
45 per cent of the Cubmasters are in satisfactory units.

The lower standing of the cub units and cubmasters is doubtless due in larger measure to two peculiarities of this comparatively new younger-boy program: (1) The cubmaster, whether trained or untrained, has much less contact than prevails in troops and ships with the boys themselves, whose chief personal relationships are with Den Chiefs (who are scouts), and Den Mothers; (2) the training program of cubbing is still in a state of flux.

(3) A third question relates to the amount of training of leaders. All relationships thus far described refer to the minimum of formal training, the elementary course. Does additional training further improve unit efficiency? Any answer to this question is complicated

by the fact that courses vary in length, quality, and relevancy to general troop efficiency; although the elementary, standard, and specialty courses each count as one course. Nevertheless, it is instructive to make a comparison between unit rating and the amount of training their leaders have had—none, one, two, three, four or more courses (Fig. 7). This relationship can be summarized by saying: thirty-one per cent of leaders with no training courses are in satisfactory units; fifty-one per cent of leaders with one course are in satisfactory units; sixty per cent of those with two courses; fifty-three per cent of those with three courses; and sixty-five per cent of those with four or more courses; while the corresponding frequencies in unsatisfactory units declines from sixty-nine per cent of men with no training to thirty-five per cent of the men with four or more courses.

Although the measured effect of additional training beyond the minimum does not appear to be extremely great, it is positive and fairly consistent. As might be expected, the largest difference is between the wholly untrained leader and the leader who has completed the minimum formal training. If the specialty courses were omitted, and the advanced Standard course alone studied in this connection, it is quite possible that the progression might be clearer and more marked. As it is, the increased frequency of satisfactory ratings with longer training indicates that the best volunteer scout leader is one who considers his job one about which he can never learn too much. Certainly, this, as well as the entire series of comparisons herein reported, offers positive evidence that the scout leadership training program, for all its limitations, does bear a definite and important relation to scouting efficiency.

THE CALCULATION OF PARTIAL AND MULTIPLE COEFFICIENTS OF REGRESSION AND CORRELATION

F. L. WREN

George Peabody College for Teachers

EDITOR'S NOTE.—This is a simple straightforward method which should be particularly valuable to the inexperienced worker. Essentially, it consists of reducing a 2×2 determinant over and over, that is, a multiplication using the plus bar followed by a multiplication using the minus bar of a calculating machine. It is possible that when the number of variables is large that the simplicity of the method will be offset by the large number of operations to be performed. J. W. D.

It is the purpose of this paper to present a systematic method for calculating partial and multiple coefficients of regression and correlation which has distinct advantages over present methods. There are no recursion formulae to be remembered such as those of Yule[1] nor elaborate determinants to expand as in the method presented by Kelley.[2] Although there is some similarity to the Doolittle method, the one presented here was developed independently as an application of a simplified technique for solving systems of linear equations,[3] and it is simpler and more direct than certain published simplifications of the Doolittle method.[4]

I. PARTIAL REGRESSION AND CORRELATION

In order to demonstrate the technique for calculating partial coefficients of correlation we shall assume a four-variable problem

[1] Yule, G. U.: *An Introduction to the Theory of Statistics.* London: Charles Griffin and Co., Ltd., (6) 1922, pp. 229–252.

[2] Kelley, T. L.: *Statistical Method.* New York: The Macmillan Co., 1923, pp. 279–302.

[3] Wren, F. L.: "Neo-Sylvester Contractions and the Solution of Systems of Linear Equations." *Bulletin of the American Mathematical Society*, Vol. XLIII, December, 1937, pp. 823–824.

[4] Peters, C. C. and Wykes, E. C.: "Simplified Method for Computing Regression Coefficients and Multiple and Partial Correlations." *Journal of Educational Research*, Vol. XXIII, June, 1931, pp. 383–393.

Thomson, G. H.: "Some Points of Mathematical Technique in the Factorial Analysis of Ability." *Journal of Educational Psychology*, Vol. XXVII, January, 1936, pp. 37–54.

Kurtz, A. K.: "The Use of the Doolittle Method in Obtaining Related Multiple Correlation Coefficients." *Psychometrika*, Vol. I, March, 1936, pp. 45–51.

and write the ordinary product-moment coefficients in the following array:

ARRAY I

	1	2	3	4	Check
1	1	.7516	.8402	.7051	3.2969
2	.7516	1	.9135	.8708	3.5339
3	.8402	.9135	1	.8319	3.5856
4	.7051	.8708	.8319	1	3.4078

In the first four columns we have all of the product-moment coefficients and, since $r_{ij} = r_{ji}$, this is a symmetric array. Each element of the *Check* column is merely the sum of the elements of its corresponding row, for example:

$$3.2969 = 1 + .7516 + .8402 + .7051.$$

Now let the problem be to determine $\beta_{ij.3}$ and $r_{ij.3}$ where i and j take the values 1, 2, 4. Block out row 3 and column 3 and use Steps 1–4 to determine the elements of Array II.

Step 1.—Multiply the element in the ith row ($i \neq 3$) and jth column ($j \neq 3$) by the element in row 3 and column 3.

Step 2.—Multiply the element in row i and column 3 by the element in row 3 and column j.

Step 3.—Subtract the product obtained in Step 2 from the product obtained in Step 1.

Step 4.—Use the result of Step 3 as the element for the ith row and jth column of Array II.

For example, the elements of row 1 of Array II are obtained as follows:

(row 1 column 1) $.2941 = (1 \times 1) - (.8402 \times .8402)$
(row 1 column 2) $-.0159 = (.7516 \times 1) - (.8402 \times .9135)$
(row 1 column 4) $.0061 = (.7051 \times 1) - (.8402 \times .8319)$
(row 1 column check) $.2843 = (3.2969 \times 1) - (.8402 \times 3.5856)$

ARRAY II

	1	2	4	Check
1	.2941	−.0159	.0061	.2843
2	−.0159	.1655	.1109	.2605
4	.0061	.1109	.3079	.4249

The calculations for the elements of Array II may be checked by the fact that each element of the Check column, although obtained independent of the elements of columns 1, 2 and 4 is the sum of the elements in its respective row. For example:

$$.2843 = .2941 - .0159 + .0061.$$

[*Note.*—Due to the rounding off of decimals there may be at times a slight difference in the last decimal place. This should be adjusted before proceeding further with the calculations.]

Furthermore, it is to be noticed that the elements of columns 1, 2 and 4 are symmetrically arranged. This fact may be used in either of two ways; namely, (1) *to reduce the number of calculations*, for example when the element for row 1 column 2 is found it also may be placed in row 2 column 1; or (2) *as a check*, for example if the element for row 1 and column 2 is calculated and then the element for row 2 column 1 is calculated independently, the two should be the same.

All regression and correlation coefficients with variable 3 controlled may be read from Array II as follows:

Step 5.—$\beta_{ij\cdot 3}$ $(i \neq j)$ is obtained by dividing the element in row i and column j by the element in row j and column j.

Step 6.—$r_{ij\cdot 3}$ $(i \neq j)$ is obtained by dividing the element in row i and column j by the square root of the product of the element in row i and column i by the element in row j and column j. For example:

$$\beta_{1\,23} = \frac{-.0159}{.1655} = -.10 \qquad \beta_{21\cdot 3} = \frac{-.0159}{.2941} = -.05$$

$$r_{12\cdot 3} = \frac{-.0159}{\sqrt{(.2941)(.1655)}} = -.07$$

$$\beta_{24\cdot 3} = \frac{.1109}{.3079} = .36 \qquad \beta_{42\cdot 3} = \frac{.1109}{.1655} = .67$$

$$r_{24\cdot 3} = \frac{.1109}{\sqrt{(.1655)(.3079)}} = .49$$

If now the problem is to find $\beta_{12\cdot 34}$, $\beta_{21\cdot 34}$ and $r_{12\cdot 34}$ we block out row 4 and column 4 in Array II and proceed as before using row 4 and

ARRAY III

	1	2	Check
1	.0906	−.0056	.0850
2	−.0056	.0387	.0331

column 4 in place of row 3 and column 3, respectively. Steps 1–4 now give us Array III.

The Check column may be used as a check on the calculations just as in the previous case. By means of Steps 5 and 6 we read the following results from Array III:

$$\beta_{12\cdot34} = \frac{-.0056}{.0387} = -.15 \qquad \beta_{21\cdot34} = \frac{-.0056}{.0906} = -.06$$

$$r_{12\cdot34} = \frac{-.0056}{\sqrt{(.0906)(.0387)}} = -.10$$

It should be noted that if the problem had been to determine $r_{14\cdot23}$, we should have blocked out row 2 and column 2 in Array II and proceeded as above. To generalize the technique, it is to be noted that the controlling of any specified variable is obtained by blocking out the row and the column whose label corresponds to the label of the variable to be controlled. Steps 1–4 give a new array of constants which may be checked by use of the Check column and from which regression and correlation coefficients may be read through the technique of Steps 5 and 6. Furthermore, the number of variables is immaterial, as each separate application of Steps 1–6 has the effect of controlling one variable; namely, the one whose row and column is blocked out before starting the calculations.

II. MULTIPLE REGRESSION AND CORRELATION

With a slight modification the same technique may be used to calculate multiple coefficients of correlation. We shall use the same product-moment coefficients of correlation and assume, for sake of definiteness that variable 1 is the criterion variable. Furthermore, we shall assume that the variables are so arranged that the following multiple coefficients are desired in the order listed:

$$R_{1\cdot2};\ R_{1\cdot23};\ R_{1\cdot234}$$

In this problem we shall start with Array IV. It is to be noted that, if the Check column is omitted, Array IV is the same as Array I except for two differences; namely,

(1) In column 1 (the criterion column) the elements are opposite in sign to those of Array I except in row 1 (the criterion row) where the element is zero.

(2) There is an added column (Column *D*) in which all of the elements are zero except in row 1 and column 1 (the criterion row and

ARRAY IV

	1	2	3	4	*D*	Check
1	0	.7516	.8402	.7051	1	3.2969
2	−.7516	1	.9135	.8708	0	2.0327
3	−.8402	.9135	1	.8319	0	1.9052
4	−.7051	.8708	.8319	1	0	1.9976

column) where the element is 1. This column will be known as the *divisor* column.

As in all previous arrays the Check Column is made up of elements each of which is the sum of the elements in its respective row.

Since the first value desired is $R_{1\cdot 2}$ we now block out row 2 and column 2 and perform the calculations of Steps 1–4, using row 2 and column 2 in place of row 3 and column 3 respectively. These calculations give the elements of Array V. It is to be noted that columns

ARRAY V

	1	3	4	*D*	Check
1	.5649	.1536	.0506	1	1.7691
3	−.1536	.1655	.0364	0	.0483
4	−.0506	.0364	.2417	0	.2275

1, 3 and 4 of Array V are symmetric except that there is a difference of sign between the elements of row 1 and those of column 1. This fact may be used, as in the previous cases, either as a check or to reduce the number of calculations. The Check column is used just as in the previous cases.

To obtain $R^2_{1\cdot 2}$ from Array V, divide the element in the criterion row (row 1) and column (column 1) by the element in the criterion row and divisor column (column *D*), that is

$$R^2_{1\cdot 2} = \frac{.5649}{1} = .5649$$

$$R_{1\cdot 2} = \sqrt{.5649} = .75$$

ARRAY VI

	1	4	*D*	Check
1	.1171	.0028	.1655	.2854
4	−.0028	.0387	0	.0359

To carry the calculations further to obtain $R_{1\cdot 23}$, block out row 3 and column 3 in Array V and repeat the calculations of Steps 1–4 to obtain Array VI, whose elements may be checked as in all previous cases. $R^2_{1\cdot 23}$ is obtained from Array VI in the same way that $R^2_{1\cdot 2}$ was obtained from Array V. Hence

$$R^2_{1\cdot 23} = \frac{.1171}{.1655} = .7076$$
$$R_{1\cdot 23} = .84$$

By blocking out row 4 and column 4 in Array VI and proceeding as before we obtain Array VII from which we find that

ARRAY VII

	1	*D*	Check
1	.004540	.006404	.010944

$$R^2_{1\cdot 234} = \frac{.004540}{.006404} = .7089$$
$$R_{1\cdot 234} = .84$$

Although in general practice it is customary to so arrange the variables that the criterion variable is that one with the leading label, that is not necessary in the present technique. Any variable may be used as the criterion variable by simply remembering the following procedure:

(1) The starting array, Array IV in the illustrative example, is obtained by writing down all the product-moment coefficients and then replacing those in the criterion column by their negatives except that a zero is written in the criterion row and column.

(2) The divisor column is obtained by placing zeros in every row except the criterion row, where a 1 is placed.

(3) After each set of calculations the corresponding R^2 is obtained by dividing the element in the criterion row and criterion column by the element in the criterion row and divisor column.

As in the case of the calculation of partial coefficients of regression and correlation this technique is not altered by increasing the number of variables. It is only necessary to remember that a variable is controlled by blocking out its corresponding row and column and carrying out the indicated calculations.

TWO METHODS OF ESTIMATING BETA WEIGHTS

ROBERT J. WHERRY

University of North Carolina

The methods to be discussed are, as the title implies, approximation methods. While the results are in neither case exact, they are nevertheless methods of obtaining quite close approximations without any of the fumbling and guessing inherent in the Kelley-Salisbury iteration method with approximately equally good results.

The two methods presented differ considerably in degree of complexity, the more complex formula yielding the better results, however. Being approximations neither of the formulae of course is susceptible to rigid mathematical proof. The simpler formula involves assumptions of equalities which do not usually exist, while the more complex formula in addition to less widespread assumptions of equality also neglects certain higher powers of decimals as being inconsequential with reference to lower powers.

The merit of these formulae lies in the fact that they offer a very quick method of finding approximations to the results which might be expected from actual calculation of the exact coefficients, thus enabling a person to: (1) decide whether the exact values are worth obtaining, (2) obtain immediately approximation answers which he might otherwise have to wait hours or days to obtain, and (3) select certain variables from a larger group which may then be given an exact solution.

THE SIMPLE METHOD: A FIRST APPROXIMATION

The theoretical derivation of this formula, if it may be so called, follows a method developed by Kurtz[1] by which he obtained an approximate value for the multiple correlation coefficient itself. The resulting formula in this case however is quite different.

H. L. Reitz[2] gives the following formula for the beta weight in terms of determinants

$$B_{ou} = -\frac{R_{ou}}{R_{oo}} \tag{1}$$

where R_{AB} stands for the determinant composed of the coefficients of the usual normal equations with row A and column B eliminated, where R is thus

$$R = \begin{vmatrix} 1.00 & r_{12} & r_{13} & \cdots & r_{1M} & r_{01} \\ r_{12} & 1.00 & r_{23} & \cdots & r_{2M} & r_{02} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{1M} & r_{2M} & r_{3M} & \cdots & 1.00 & r_{0M} \\ r_{01} & r_{02} & r_{03} & & r_{0M} & 1.00 \end{vmatrix} \tag{2}$$

If now we assume that each of the criterion correlations is equal to its average value, r_0, and that in like manner each of the intercorrelations is equal to its average value, r_1, we may write the beta equation in the form given below:

$$B_{ou} = \frac{\begin{vmatrix} 1.00 & r_1 & r_1 & \cdots & r_1 & r_0 \\ r_1 & r_1 & r_1 & \cdots & r_1 & r_0 \\ r_1 & 1.00 & r_1 & \cdots & r_1 & r_0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_1 & r_1 & \cdots & r_1 & r_0 \\ r_1 & r_1 & r_1 & \cdots & 1.00 & r_0 \end{vmatrix}}{\begin{vmatrix} 1.00 & r_1 & r_1 & \cdots & r_1 & r_1 \\ r_1 & 1.00 & r_1 & \cdots & r_1 & r_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_1 & r_1 & \cdots & 1.00 & r_1 \\ r_1 & r_1 & r_1 & \cdots & r_1 & 1.00 \end{vmatrix}} \tag{3}$$

The second variable was eliminated it being immaterial which it was.

If now we let $c = r_0/r_1$, and subtract c times the first column from the last column, and one times the first column from the remaining columns in the numerator; and if in the denominator we subtract the last column from each of the other columns we obtain:

$$B_{ou} = \frac{\begin{vmatrix} 1.00 & r_1 - 1 & r_1 - 1 & \cdots & r_1 - 1 & r_0 - c \\ r_1 & 0 & 0 & \cdots & 0 & 0 \\ r_1 & 1 - r_1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & 0 & 0 & \cdots & 0 & 0 \\ r_1 & 0 & 0 & \cdots & 1 - r_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 - r_1 & 0 & 0 & \cdots & 0 & r_1 \\ 0 & 1 - r_1 & 0 & \cdots & 0 & r_1 \\ 0 & 0 & 1 - r_1 & \cdots & 0 & r_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 - r_1 & r_1 \\ r_1 - 1 & r_1 - 1 & r_1 - 1 & \cdots & r_1 - 1 & 1.00 \end{vmatrix}} \tag{4}$$

If now we add all other rows to the last row in the denominator we obtain:

$$B_{ou} = \frac{-\begin{vmatrix} 1.00 & r_1 - 1 & r_1 - 1 & \dots & r_1 - 1 & r_0 - c \\ r_1 & 0 & 0 & \dots & 0 & 0 \\ r_1 & 1 - r_1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_1 & 0 & 0 & \dots & 0 & 0 \\ r_1 & 0 & 0 & \dots & 1 - r_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 - r_1 & 0 & 0 & \dots & 0 & r_1 \\ 0 & 1 - r_1 & 0 & \dots & 0 & r_1 \\ 0 & 0 & 1 - r_1 & \dots & 0 & r_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 - r_1 & r_1 \\ 0 & 0 & 0 & \dots & 0 & 1 + (M - 1)r_1 \end{vmatrix}} \tag{5}$$

which when expanded yields the equation:

$$B_{ou} = \frac{-\left(r_0 - \frac{r_0}{r_1}\right)(1 - r_1)^{M-2} r_1}{|1 + (M - 1)r_1|(1 - r_1)^{M-1}} = \frac{r_0}{1 + (M - 1)r_1} \tag{6}$$

which is the value that would be assigned to the betas if all betas were assumed to be equal. By a little consideration of the formula (6) it is apparent that certain substitutions should be made. In the first place it is obvious that the individual weights are much more influenced by their own criterion correlations than they are by the average or by any one of the others, thus as an approximation to the true weights we may replace r_0 by r_{0u}. It is also obvious that the intercorrelations of the variable under consideration (u) have more to do with that particular weight than do the remaining intercorrelations; thus in place of r_1 we substitute $\bar{r}_{ux}$. Our formula thus changed becomes:

$$B_{ou} = \frac{r_{ou}}{1 + (M - 1)\frac{\Sigma_1^M r_{ux}}{M - 1} x \neq u} = \frac{r_{ou}}{1 + \Sigma_1^M r_{uxx \neq u}} \tag{7}$$

In order to illustrate the use of the formula and the ease with which it may be applied, the following problem involving a criterion and six independent variables, as given in Table I below, will be solved:

$$R = \begin{vmatrix} 1.00 & r_{12} & r_{13} & \cdots & r_{1M} & r_{01} \\ r_{12} & 1.00 & r_{23} & \cdots & r_{2M} & r_{02} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{1M} & r_{2M} & r_{3M} & \cdots & 1.00 & r_{0M} \\ r_{01} & r_{02} & r_{03} & & r_{0M} & 1.00 \end{vmatrix} \tag{2}$$

If now we assume that each of the criterion correlations is equal to its average value, r_0, and that in like manner each of the intercorrelations is equal to its average value, r_1, we may write the beta equation in the form given below:

$$B_{ou} = \frac{\begin{vmatrix} 1.00 & r_1 & r_1 & \cdots & r_1 & r_0 \\ r_1 & r_1 & r_1 & \cdots & r_1 & r_0 \\ r_1 & 1.00 & r_1 & \cdots & r_1 & r_0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_1 & r_1 & \cdots & r_1 & r_0 \\ r_1 & r_1 & r_1 & \cdots & 1.00 & r_0 \end{vmatrix}}{\begin{vmatrix} 1.00 & r_1 & r_1 & \cdots & r_1 & r_1 \\ r_1 & 1.00 & r_1 & \cdots & r_1 & r_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & r_1 & r_1 & \cdots & 1.00 & r_1 \\ r_1 & r_1 & r_1 & \cdots & r_1 & 1.00 \end{vmatrix}} \tag{3}$$

The second variable was eliminated it being immaterial which it was.

If now we let $c = r_0/r_1$, and subtract c times the first column from the last column, and one times the first column from the remaining columns in the numerator; and if in the denominator we subtract the last column from each of the other columns we obtain:

$$B_{ou} = \frac{\begin{vmatrix} 1.00 & r_1 - 1 & r_1 - 1 & \cdots & r_1 - 1 & r_0 - c \\ r_1 & 0 & 0 & \cdots & 0 & 0 \\ r_1 & 1 - r_1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_1 & 0 & 0 & \cdots & 0 & 0 \\ r_1 & 0 & 0 & \cdots & 1 - r_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 - r_1 & 0 & 0 & \cdots & 0 & r_1 \\ 0 & 1 - r_1 & 0 & \cdots & 0 & r_1 \\ 0 & 0 & 1 - r_1 & \cdots & 0 & r_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 - r_1 & r_1 \\ r_1 - 1 & r_1 - 1 & r_1 - 1 & \cdots & r_1 - 1 & 1.00 \end{vmatrix}} \tag{4}$$

If now we add all other rows to the last row in the denominator we obtain:

$$B_{ou} = \frac{-\begin{vmatrix} 1.00 & r_1 - 1 & r_1 - 1 & \dots & r_1 - 1 & r_0 - c \\ r_1 & 0 & 0 & \dots & 0 & 0 \\ r_1 & 1 - r_1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_1 & 0 & 0 & \dots & 0 & 0 \\ r_1 & 0 & 0 & \dots & 1 - r_1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 - r_1 & 0 & 0 & \dots & 0 & r_1 \\ 0 & 1 - r_1 & 0 & \dots & 0 & r_1 \\ 0 & 0 & 1 - r_1 & \dots & 0 & r_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 - r_1 & r_1 \\ 0 & 0 & 0 & \dots & 0 & 1 + (M - 1)r_1 \end{vmatrix}} \tag{5}$$

which when expanded yields the equation:

$$B_{ou} = \frac{-\left(r_0 - \frac{r_0}{r_1}\right)(1 - r_1)^{M-2} r_1}{|1 + (M - 1)r_1|(1 - r_1)^{M-1}} = \frac{r_0}{1 + (M - 1)r_1} \tag{6}$$

which is the value that would be assigned to the betas if all betas were assumed to be equal. By a little consideration of the formula (6) it is apparent that certain substitutions should be made. In the first place it is obvious that the individual weights are much more influenced by their own criterion correlations than they are by the average or by any one of the others, thus as an approximation to the true weights we may replace r_0 by r_{0u}. It is also obvious that the intercorrelations of the variable under consideration (u) have more to do with that particular weight than do the remaining intercorrelations; thus in place of r_1 we substitute $\bar{r}_{ux}$. Our formula thus changed becomes:

$$B_{ou} = \frac{r_{ou}}{1 + (M - 1)\frac{\Sigma_1^M r_{ux}}{M - 1} x \neq u} = \frac{r_{ou}}{1 + \Sigma_1^M r_{uxx \neq u}} \tag{7}$$

In order to illustrate the use of the formula and the ease with which it may be applied, the following problem involving a criterion and six independent variables, as given in Table I below, will be solved:

TABLE I.—CRITERION AND INTERCORRELATIONS (ORIGINAL SIGNS)

Variable	1	2	3	4	5	6	0
1	1.0000	−.0388	.1415	.1322	.3407	.2476	−.1676
2	−.0388	1.0000	.0053	−.0530	−.3104	−.1202	.3392
3	.1415	.0053	1.0000	−.1553	−.0849	.0344	−.2152
4	.1322	−.0530	−.1553	1.0000	.0479	−.1562	.1826
5	.3407	−.3104	−.0849	.0479	1.0000	.0503	−.3038
6	.2476	−.1202	.0344	−.1562	.0503	1.0000	−.1913
0	−.1676	.3392	−.2152	.1826	−.3038	−.1913	1.0000

Since the true influence of the intercorrelations is due not to their absolute value but to their signs as well, it is necessary to have all criterion correlations positive in sign before formula (7) is applied. We must, therefore, change the signs of both row and column for variables 1, 3, 5, and 6. (If these changes are made with a red pencil the table will not have to be recopied). Performing these changes we obtain:

TABLE II.—CRITERION AND INTERCORRELATIONS (ALL CRITERION CORRELATIONS POSITIVE)

Variable	1	2	3	4	5	6	0
1	1.0000	.0388	.1415	−.1322	.3407	.2476	.1676
2	.0388	1.0000	−.0053	−.0530	.3104	.1202	.3392
3	.1415	−.0053	1.0000	.1553	−.0849	.0344	.2151
4	−.1322	−.0530	.1553	1.0000	−.0479	.1562	.1826
5	.3407	.3104	−.0849	−.0479	1.0000	.0503	.3038
6	.2476	.1202	.0344	.1562	.0503	1.0000	.1913
0	.1676	.3392	.2151	.1826	.3038	.1913	1.0000
Σ_1^6	1.6364	1.4111	1.2410	1.0784	1.5638	1.6087	

Now if one examines the method of obtaining the columnar sums at the foot of Table II, it is seen that these values are equal to $1 + \Sigma_1^M r_{UX_{U \neq X}}$, which is the denominator of equation (7). Thus to obtain our approximate beta weights we have only to divide the entries in the criterion column, with signs as in Table I, by the corresponding columnar sums of Table II. For this particular problem we have:

TABLE III.—APPROXIMATE BETA WEIGHTS

U	1	2	3	4	5	6
B_{0U}.....	$\frac{-.1676}{1.6364}$	$\frac{.3392}{1.4111}$	$\frac{-.2151}{1.2410}$	$\frac{.1826}{1.0784}$	$\frac{-.3038}{1.5688}$	$\frac{-.1913}{1.6087}$

Comparison of these values with the true beta weights derived by the Doolittle process, and with the resulting multiple correlation coefficients for the two sets of weights gives:

TABLE IV.—APPROXIMATE AND ACTUAL COEFFICIENTS

	B_{01}	B_{02}	B_{03}	B_{04}	B_{05}	B_{06}	$R_{0\cdot12\ldots6}$
Approximations...	−.1024	.2404	−.1734	.1693	−.1937	−.1189	.5008
Actual.....	−.0492	.2651	−.1991	.1668	−.2244	−.1031	.5093

In Table V below are presented the results obtained in this and six other problems which range from six to ten independent variables.

TABLE V.—ERRORS IN MULTIPLE R RESULTING FROM THE USE OF THE WHERRY SIMPLE APPROXIMATION BETAS

Variables	Predicted R	Actual R	Error
6	.5008	.5093	−.0085
6	.4175	.4961	−.0830
6	.4379	.4431	−.0052
8	.3768	.4247	−.0479
8	.4099	.4465	−.0366
10	.4175	.4961	−.0786
10	.6029	.6281	−.0252
10	.7279	.7869	−.0590

The mean error of these eight examples together with their average deviation is only

$$M_{error} = .0430 \pm .0241 \tag{8}$$

THE COMPLEX METHOD: A SECOND APPROXIMATION

The formula for the second approximation method grows directly out of the literal expansion of the betas for 2, 3, 4, etc., independent variables. Thus we have:

$$B_{01.2} = \frac{r_{01} - r_{02}r_{12}}{1 - r^2_{12}} \tag{9}$$

$$B_{01.23} = \frac{\begin{matrix} - r_{02}r_{12} & + r_{02}r_{13}r_{23} \\ r_{01} - r_{03}r_{13} - r_{01}r^2_{23} & + r_{03}r_{12}r_{23} \end{matrix}}{\begin{matrix} 1 & - r^2_{12} \\ & - r^2_{13} \end{matrix} - r^2_{23} + 2r_{12}r_{13}r_{23}} \tag{10}$$

$$B_{01.234} = \frac{\begin{matrix} & & + r_{02}r_{14}r_{24} \\ & & + r_{02}r_{13}r_{23} \\ - r_{02}r_{12} & - r_{01}r^2_{23} & + r_{03}r_{12}r_{23} \\ r_{01} - r_{03}r_{13} & - r_{01}r^2_{34} & + r_{03}r_{14}r_{24} \\ - r_{04}r_{14} & - r_{01}r^2_{24} & + r_{04}r_{12}r_{24} \\ & & + r_{04}r_{13}r_{34} \end{matrix} + k(r_{xx})^4}{\begin{matrix} - r^2_{12} - r^2_{23} + 2r_{12}r_{13}r_{23} \\ 1 - r^2_{13} - r^2_{34} + 2r_{13}r_{14}r_{34} + 2r_{23}r_{24}r_{34} + k(r_{xx})^4 \\ - r^2_{14} - r^2_{24} + 2r_{12}r_{14}r_{24} \end{matrix}} \tag{11}$$

TABLE VI.—MULTIPLIERS FOR FORMULA (12)

n	$(n-1)$	$\frac{(n-1)(n-2)}{2}$	$(n-1)(n-2)$	$\frac{(n-1)(n-2)(n-3)}{3}$
2	1	0	0	0
3	2	1	2	0
4	3	3	6	2
5	4	6	12	8
6	5	10	20	20
7	6	15	30	40
8	7	21	42	70
9	8	28	56	112
10	9	36	72	168
11	10	45	90	240
12	11	55	110	330
13	12	66	132	440
14	13	78	156	572
15	14	91	182	728
16	15	105	210	906
17	16	120	240	1120
18	17	136	272	1360
19	18	153	306	1632
20	19	171	342	1938

Whence, disregarding all terms where four or more r's are multiplied together, we obtain the expression:

$$B_{ov} = \frac{r_{ov}\left[1 - \frac{(n-1)(n-2)}{2}\bar{r}^2_{pq}\right] - \bar{r}_{op}\bar{r}_{Up}[(n-1) - (n-1)(n-2)\bar{r}_{pq}]}{\left[1 - \frac{(n-1)(n-2)}{2}\bar{r}^2_{pq} - \bar{r}^2_{Up}((n-1) - (n-1)(n-2)\bar{r}_{pq}) + \frac{(n-1)(n-2)(n-3)}{3}\bar{r}^3_{pq}\right]} \quad (12)$$

which looks terrifying, but which is fairly easy to apply in practice, since each of the individual betas is obtained in exactly the same manner and since a large part of the denominator involves the same terms as the numerator. The multipliers for the averages up through twenty variables may be obtained from Table VI.

Thus, for the problem worked out in the first part of the paper (six independent variables) equation (12) becomes:

$$B_{ov} = \frac{r_{ov}(1 - 10\bar{r}^2_{pq}) - \bar{r}_{op}\bar{r}_{Up}(5 - 20\bar{r}_{pq})}{1 - 10\bar{r}^2_{pq} - \bar{r}^2_{Up}(5 - 20\bar{r}_{pq}) + 20\bar{r}^3_{pq}} \quad (13)$$

From Table II we find the following values:

U	r_{ov}	$\bar{r}_{op}$	$\bar{r}_{Up}$	$\bar{r}_{pq}$
1	.1676	.2464	.1273	.0763
2	.3392	.2121	.0822	.0853
3	.2151	.2369	.0482	.0921
4	.1826	.2434	.0157	.0986
5	.3038	.2192	.1138	.0790
6	.1913	.2417	.1217	.0774

$$\Sigma r_{ov} = 1.3996 \qquad \Sigma\bar{r}_{Up} = .5089$$

where

$$\bar{r}_{op} = \frac{\Sigma r_{ov} - r_{ov}}{n-1}; \qquad \bar{r}_{Up} = \frac{\text{Columnar Sum}_{\text{Table II}^{-1}}}{n-1};$$

$$\bar{r}_{pq} = \frac{\Sigma\bar{r}_{Up} - \bar{r}_{Up}}{n-1}$$

whence, prefixing the signs from Table I, we have

$$B_{01} = -\frac{.1676|1 - 10(.0763)^2| - (.2464)(.1273)|5 - 20(.0763)|}{1 - 10(.0763)^2 - (.1273)^2|5 - 20(.0763)| + 20(.0763)^3} = -.0545 \quad (14)$$

$$B_{02} = \frac{.3392|1 - 10(.0853)^2| - (.2121)(.0822)|5 - 20(.0853)|}{1 - 10(.0853)^2 - (.0822)^2|5 - 20(.0853)| + 20(.0853)^3} = .2784 \quad (15)$$

$$B_{03} = -\frac{.2151|1 - 10(.0921)^2| - (.2369)(.0482)|5 - 20(.0921)|}{1 - 10(.0921)^2 - (.0482)^2|5 - 20(.0921)| + 20(.0921)^3} = -.1742 \quad (16)$$

$$B_{04} = \frac{.1826|1 - 10(.0986)^2| - (.2434)(.0157)|5 - 20(.0986)|}{1 - 10(.0986)^2 - (.0157)^2|5 - 20(.0986)| + 20(.0986)^3} = .1678 \quad (17)$$

$$B_{05} = -\frac{.3038|1 - 10(.0790)^2| - (.2192)(.1138)|5 - 20(.0790)|}{1 - 10(.0790)^2 - (.1138)^2|5 - 20(.0790)| + 20(.0790)^3} = -.2211 \quad (18)$$

$$B_{06} = -\frac{.1913|1 - 10(.0774)^2| - (.2417)(.1217)|5 - 20(.0774)|}{1 - 10(.0774)^2 - (.1217)^2|5 - 20(.0774)| + 20(.0774)^3} = -.0872 \quad (19)$$

which, when compared to the true beta weights and the true multiple R give the results in Table VII.

TABLE VII.—APPROXIMATE AND ACTUAL COEFFICIENTS

	B_{01}	B_{02}	B_{03}	B_{04}	B_{05}	B_{06}	$R_{0.12\ldots6}$
Approximations . .	−.0545	.2784	−.1742	.1678	−.2211	−.0872	.5087
Actual	−.0492	.2651	−.1991	.1668	−.2244	−.1031	.5093

The writer has also applied this formula to one of the ten variable problems reported in the earlier part of the paper. Table VIII gives a comparison of the results obtained by means of the Doolittle and each of the two approximation equations for the six and ten variable problems.

TABLE VIII.—MULTIPLE R OBTAINED BY USE OF BETAS OBTAINED BY VARIOUS METHODS

M	Wherry first approximation equation 7	Wherry second approximation equation 12	Doolittle
6	.5008	.5087	.5093
10	.6029	.6137	.6281

The multiple (combination) correlation coefficient is most easily obtained and systematically calculated from approximation weights

by use of the form used in the Kelley-Salisbury iteration method. If one wishes one can then proceed by the usual method of correcting these weights to achieve any desired degree of accuracy.

A further step in improving the weights is that of multiplying all of them by a constant which makes the mean of the predicted scores equal to the mean of the actual scores. Thus, if R be found by the Kelley-Salisbury method as suggested, and using the value of the problem above as an illustration, we have:

$$R = \frac{S_c}{\sqrt{\sigma_c{}^2}} = \frac{.2557}{\sqrt{.2526}} \tag{20}$$

The constant multiplier for the weights is found by the relationship

$$C = \frac{S_c}{\sigma_c{}^2} = \frac{.2557}{.2526} = 1.0123 \tag{21}$$

Since each of the weights is multiplied by this same quantity, it, of course, has no influence upon the correlation coefficient. If the beta weights found by equation (12) are multiplied by the C of equation (21), we have

$$\begin{aligned} B_{01} &= -.0552 & B_{04} &= .1699 \\ B_{02} &= .2818 & B_{05} &= -.2238 \\ B_{03} &= -.1763 & B_{06} &= -.0883 \end{aligned}$$

The readers attention is also called to an article by H. A. Edgerton,[3] who has worked out a method of obtaining average intercorrelation coefficients when the correlations themselves have not yet been worked out, thus making it possible to still use this method even though a table of intercorrelations is not available.

BIBLIOGRAPHY

1. Kurtz, A. K.: "A Special Case of the Multiple Correlation Coefficient." *Jour. of Educ. Psychol.*, Vol. xx, 1929, pp. 378–385.
2. Reitz, H. L.: *Mathematical Statistics.* Chicago: Open Court Publishing Company, 1927, p. 5.
3. Edgerton, H. A.: "A Formula for Finding the Average Correlation of Any One Variable with the $(n-1)$ Other Variables Without Solving Any of the Individual Correlations." *Jour. of Educ. Psychol.*, Vol. xxvi, 1935, pp. 373–376.

BOOK REVIEW

ERNEST HORN. *Methods of Instruction in the Social Studies.* Part XV of the Report of the Commission on the Social Studies of the American Historical Association. New York: Charles Scribners' Sons, 1937.

This volume is not a manual of specific practices for social science teachers, but a discussion of the fundamental problems of learning and instruction. Although the treatment of underlying issues and principles is specifically related to methods of teaching in the social studies, the book is in fact a contribution of first importance to educational psychology.

The volume is one of the most thoroughly documentated treatises in the educational literature. It is by no means, however, merely a collection of summaries of researches and theoretical discussions. Before writing, Horn made an exhaustive and critical analysis of the experimental literature, not only on social studies instruction, but also on the fundamental problems of learning. He then focused the useful results of this research, together with the insights derived from his own rich experience, upon the basic problems of learning and teaching in the social studies.

The opening chapter on the "Perspective in Which Method Is to Be Viewed" points out that significant outcomes cannot result from "learning through petty teaching devices," (p. 23) but only from methods which are "instrumental" and which are "chosen and appraised in view of the ends to be reached and in the light of conditioning circumstances" (p. 38). This is a refreshing doctrine in the midst of widely acclaimed potential panaceas of instruction such as the unit method, which often become "a cult, an empty and ritualistic exercise, without vitality and without significance" (p. 38). One of the most valuable parts of the chapter is that which shows how methods of learning and instruction in the social studies should be related to methods of scholarly thinking and investigation in those disciplines.

Since instruction must be based upon a realistic knowledge of the school population, Chapter II is devoted to "Identification and Classification of Students." The problem of individual differences is presented by evidence concerning the great variability of intelligence, reading ability, and school achievement within any one school grade,

and the striking overlap in pupil characteristics from one grade to another, so that about the "only thing that can be said with assurance about a third-grade child is that he is in the third grade" (p. 42).

There are those who, because they are impressed by the dependence of action upon emotion, place little faith in an educational program which emphasizes rational bases of belief and conduct. They may have chosen another ideology from that of the totalitarian states, but they propose the adoption of the educational methods of the dictators; namely, authoritative indoctrination and emotional conditioning. It is encouraging, therefore, to find Horn calling for a greater intellectualization of the controversial issues which should be included in a dynamic program of social education.

Perhaps, the most important chapters are IV and V which deal with the problems of meaning and reading in relation to learning. The problem of meaning is probably the most significant one in the psychology of learning, but well-organized instructional material on this topic has not been available. It is probable that Horn's treatment will become required reading in most courses in educational psychology.

The presentation is essentially that of the rôle of language in the learning process. Instruction deals more often than not with symbols of reality rather than reality itself. Language is not merely a pedagogical tool; "it is not only the chief instrument of thought but also the principal means of communication" (p. 127). These two chapters are attacks upon the wide prevalence of verbalism which results from the child's inability to understand written and spoken language and his recourse to memorizing or paraphrasing the lecture or the text. The discussion is replete with illustrations of the fuzzy and erroneous meanings which children get from verbal presentations. Thus the pupil who declared in an examination that "The French Revolution wrote insulting letters to the American Revolution" justified his statement by the instructor's comment that "The French Revolution corresponded in a rough way with the American Revolution" (p. 142).

Acquiring meaning is an active process of problem-solving—constructing concepts from the learner's own experience in response to patterns of verbal stimuli. In the last analysis, meaning depends upon the learner's experience, upon the "orderliness" or organization, as well as the richness of this experience. Horn has long insisted that the central processes—those which eventuate in meaning—rather than the peripheral aspects (which have been investigated through eye movements and "trained" by phonics and mechanical drills) are

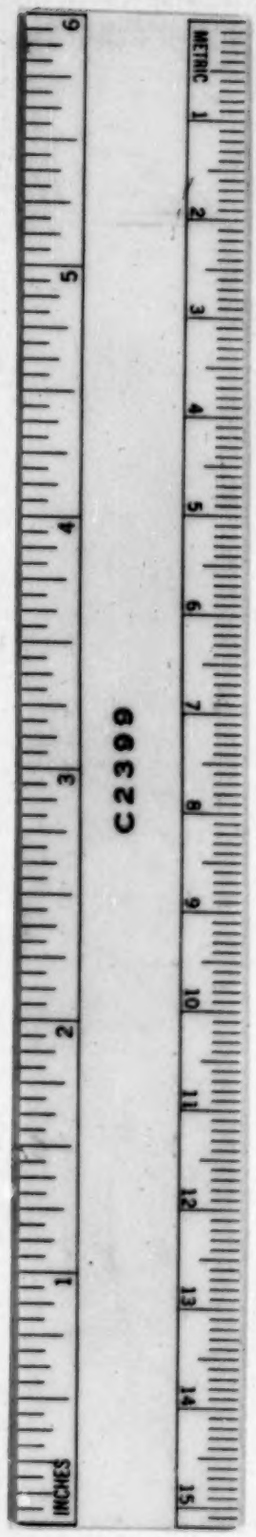
the most significant in learning to read. He points out here that little improvement in reading social studies material can be "expected from formal drill in reading unless at the same time provision is made for the enrichment of experience, the development of language abilities, and the improvement of thinking" (p. 156). The remainder of his presentation of the reading problem is devoted to an exposition of this thesis.

The treatment of "Visual Aids" (Chapter IX) and "Sources of Concrete Experience" (Chapter X) is from the point of view of enriching and extending the pupil's experience by supplementing the resources of his immediate environment. There is a valuable discussion of the use of visual aids, including a summary of experience and evidence on the usefulness of motion pictures. The author warns that each visual aid, or form of concrete experience, such as a project, must have authenticity, and that each activity should be carefully integrated with the rest of the instructional program, and used as a means to well-defined instructional outcomes. He indicts many so-called activity units as being "largely fanciful and almost wholly erroneous, or trivial and actually removed from reality." He goes on to describe types of activities which may contribute to sound learning, and outlines standards for evaluation. But he insists that constructive activities constitute a method, not *the* method of instruction.

There are several other chapters which cannot be commented upon: "The Textbook and Collateral Reading," "The Use of Imaginative Literature," "Oral Instruction," (by far the best discussion available), and "Devices for Stimulating the Imagination." The final chapter, "Provision for Future Use," summarizes the evidence on retention and applies it to learning and instruction in the social studies.

T. R. McConnell.

University of Minnesota.


INCHES
METRIC
C2309